DOCUMENT RESUME

ED 332 551 FL 019 226

AUTHOR Farrell, Paul
TITLE Vocabulary in ESP: A Lexical Analysis of the English of Electronics and a Study of Semi-Technical Vocabulary. CLCS Occasional Paper No. 25.
INSTITUTION Trinity Coll., Dublin (Ireland). Centre for Language and Communication Studies.
REPORT NO ISSN-0332-3889
PUB DATE 90
NOTE 87p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC04 Plus Postage.
DESCRIPTORS *Electronics; *English for Science and Technology; *English for Special Purposes; Foreign Countries; Language Research; Language Teachers; Postsecondary Education; Second Language Instruction; Teacher Role; Technical Writing; *Vocabulary

ABSTRACT

This report aims to show how an emphasis on vocabulary can contribute to courses in English for Specific Purposes (ESP), using the data from two lexical studies. The first study attempts to discover to what extent there may be a general language of science, or semi-technical vocabulary. This type of vocabulary would seem to offer a useful organizing principle for a lexical syllabus in common core ESP courses, and a role for the ESP teacher who finds difficulty in dealing with technical material. The study is concerned with elaborating a definition of the term semi-technical vocabulary using a combination of subjective and objective data, and by applying these criteria to the word list from the LOB Corpus study of a number of scientific disciplines. The second study examines lexical characteristics of one special language, electronics English. It is a lexical needs analysis, attempting to give a group of special interest learners words they need. The study is a count of the frequency items in 20,000 words of electronics English in ten 2,000-word samples. The entire corpus is examined, but the study also focuses on the semi-technical vocabulary of electronics, comparing it with the list developed from the first study. (JL)

TRINITY COLLEGE DUBLIN

Centre for Language and Communication Studies

Vocabulary in ESP

A lexical analysis of the English of electronics and a study of semi-technical vocabulary

Paul Farrell

CLCS Occasional Paper No.25

ISSN 0332 3889

Autumn 1990

CLCS Occasional Papers

General Editor: D. G. Little

CLCS Occasional Papers report on research carried out within or in association with the Centre for Language and Communication Studies, and on research carried out elsewhere which is of special interest to the Centre's own concerns and activities. In some instances they present the texts of public lectures promoted by the Centre.

The following titles have been published and are available from The Secretary, Centre for Language and Communication Studies, Trinity College, Dublin 2.

Autumn 1981

1. D. M. Singleton, *Language transfer: a review of some recent research* (31pp.)
2. Jeffrey L. Kallen, *Linguistics and oral tradition: the structural study of the riddle* (33pp.)
3. D. M. Singleton, *Age as a factor in second language acquisition* (70pp.) - OUT OF PRINT

Summer 1982

4. Francis Nolan, *Voice quality and speech synthesis* (30pp.)
5. Roger Bennett, *Language elicitation procedures* (25pp.)

Winter 1982-3

6. S. M. Devitt, D. G. Little, S. P. O Conchúir & D. M. Singleton, *Learning Irish with Anois is Arís* (128pp.) - OUT OF PRINT

Autumn 1983

7. William T. Littlewood, *A communicative approach to language-teaching methodology* (19pp.)
8. Rose Maclaran, *On the interaction of semantics and pragmatics* (18pp.)
9. E. M. Harding, *Compensation strategies* (54pp.)

Autumn 1984

10. Jeffrey L. Kallen, *Generative phonology in the clinic* (38pp.)
11. Gearóid O Ciaráin, *The affective dimension in second/foreign language learning: an interactional perspective* (42pp.)

Spring 1985

12. D. M. Singleton & D. G. Little, *Foreign languages at second level: defining a syllabus, with particular reference to the needs of the senior cycle* (19pp.)
13. John Harris, *The polylectal grammar stops here* (12pp.)

(Continued on inside back cover)

CLCS Occasional Paper No.25

Autumn 1990

Vocabulary in ESP: a lexical analysis of the English of electronics and a study of semi technical vocabulary*

by

Paul Farrell

0 Introduction

This paper aims to show how an emphasis on vocabulary can contribute to courses in English for Specific Purposes, using the data from two lexical studies. The first study attempts to discover to what extent there may be a general language of science, or semi-technical vocabulary. This type of vocabulary would seem to offer a useful organising principle for a lexical syllabus in common core ESP courses, and a role for the ESP teacher who finds difficulty in dealing with technical material. The study is concerned to elaborate a definition of the term semi-technical vocabulary using a combination of subjective and objective criteria, and by applying these criteria to the word list from the LOB Corpus study of a number of scientific disciplines (Hofland & Johansson 1982), draw up a list of the most prominent semi-technical items. In section 4 there is some investigation of how this list might be exploited in syllabus and materials design for common core ESP courses.

The second study is an examination of the lexical characteristics of one special language, electronics English. In simple terms, this is a lexical needs analysis; an attempt to give a group of special interest learners the words that they need. The study is a count of the frequency of occurrence of items (and other lexical data; see below) in 20,000

---

* This paper is a revised and somewhat abbreviated version of a dissertation submitted in September 1989 in fulfilment of the requirements of the M.Phil. in Applied Linguistics, Trinity College, Dublin.

words of electronics English in ten 2,000-word samples. The entire corpus is examined, but the study also focuses specifically on the semi-technical vocabulary of electronics, comparing it with the list developed from the first study.

The analysis of the English of electronics follows in the tradition of register analysis in ESP, which aims to identify grammatical and lexical features of special languages and use this information in developing materials (Hutchinson & Waters 1987, pp.9f.). Hutchinson & Waters refer to register analysis in ESP as a "stage" which occurred in the 1960s and 1970s, and express little interest in it: "... register analysis revealed that there was very little that was distinctive in the sentence grammar of Scientific English ..." (*ibid.*, p.10). Here they seem to be overlooking the lexical aspects of register, which may be far more relevant, as suggested by Sager, Dungworth & McDonald: "The non-specialist reader recognises the grammatical forms and structure of an SE (Special English) text as those normally used in written English. Many of the lexical items which he encounters there, however, are likely to be either completely new to him or at least used with a meaning different from that with which he has previously been familiar" (1980, p.230). Similarly, Johansson quotes Halliday: "the crucial criteria of any given register are to be found in its grammar and its lexis. Probably lexical features are the most obvious .... Purely grammatical distinctions between the different registers are less striking ..." (Johansson 1975, p.1).

The weaknesses of frequency studies and word lists must be freely acknowledged at the outset. The results of such studies, for example, usually show a large number of words which are relatively rare, a point made by Richards: "Statistical studies of word distribution show that ... a high percentage of words occur once only" (1974, p.72). In the present study 508 lemmas (see section 3 below for a discussion of lemmatization criteria) were found to occur more than 5 times and accounted for 18,000 words (i.e. tokens), or 90% of the corpus, but a greater number of lemmas, 750, occurred less than 5 times (many of these lemmas representing only one occurrence of a graphic word), and accounted for the remaining 10% (see section 3 below for definitions of the terms lemma, token and graphic word). As Carroll, Davies & Richman put it: "This admixture of large numbers of common words with large numbers of rare words presents a kind of paradox that is the plague of the English teacher" (1971, p.xxviii). In addition, a list of graphic words, or strings of characters, gives limited information: "... if *fly* is a frequent word, which particular meanings of the form *fly* are

frequent and in what proportions?" (Carter 1987, p.183). This is a disadvantage of the LOB Corpus data, which is a list of graphic words. In the study of electronics English however, a wealth of syntactic and semantic detail was recorded. Space limitations dictated the exclusion of this detail from this paper, but it may be examined in Farrell 1989, pp.132-202.

Another difficulty is that the data generated from such studies represents a very considerable abstraction from the original text samples. As Rossner puts it, "... language [is] inextricably bound up with context of situation, and therefore quicksilver, highly dynamic and unpredictable ... all that the lexicographers can offer the user is a selection of definitions frozen in time and incapable of admitting the vast range of uses and meanings that are likely to attach themselves to lexical items in the vast range of discourse types, situations and contexts in which they will occur" (Rossner 1985, pp.95f.). One might validly query the relevance of such text dissection and ask: "Why not give the learners the text itself and help them to understand it?" The answer to this question is that this approach is not intended to replace using authentic general or subject-specific texts to develop reading skills, but rather to supplement this by giving learners lexically-specific exercises and materials, based on empirical analysis of the special English they need to study. The data from lexical studies will enable us to give prominence in materials to lexical items which learners are likely to encounter frequently, and also to some of their common syntactic, semantic and collocational patterns. Carter notes that lexical studies of this type "offer invaluable data for vocabulary materials development" (1987, p.182).

Finally, there is the question of the extent to which the results of a survey of 20,000 words of text may be truly representative of the lexical characteristics of electronics English in general. The sample is arguably small, but ideal data would only be obtained from an ideal corpus; all texts ever published in electronics English. Any survey will reflect idiosyncrasies of the particular text samples chosen. There is some evidence that a smaller sample may be required in surveys of special languages than in the case of general language: "Statistical data can confirm that special languages have a higher rate of repetition of lexical items than general language" (Sager et al. 1980, p.238). Sager et al. go on to quote Alekseev, who in a survey of 35,000 words of electronics English "established a coverage of 98% for the 5,000 most frequent words, whereas 10,000 words are required to reach a coverage of 95% in general language" (*ibid.*). This fact is a strong argument for a

lexically-specific approach in ESP.

1 The role of the ESP teacher

Since the early 1960s there has been a rapid growth internationally in the demand for English language courses to cater for the special needs of students of scientific, technological and academic subjects, and a consequent demand for a new breed of teacher; the teacher of English for Specific Purposes. In his introduction to *Episodes in ESP*, Swales illustrates this growth by noting the increase in the number of items held in his ESP Reference Collection, from 30 in 1972 to 1300 in 1984 (Swales 1985, p.x). In spite of the obvious interest Swales and a few other writers have taken in ESP, in general teacher trainers, applied linguists and others to whom the ESP teacher might look for guidance have not kept pace with the mushrooming ESP industry, as is evidenced by the paucity of books available on the subject. Hutchinson & Waters note that "the great majority of ESP teachers have not been trained as such" (Hutchinson & Waters 1987, p.157). It is not surprising, therefore, to find that English language teachers recruited to work on ESP programmes are sometimes vague about their role. Those who have recruited them, in many cases managers of commercial enterprises, may not be helpful in clarifying their role, having simply identified a language problem and assuming that the English teacher will know how to go about remedying it. Subject teachers working alongside the English teacher may be somewhat mystified about the function of the ESP teacher. Several typical responses to this confusion, whether elaborated as a policy at senior level or simply adopted as a strategy by individual teachers, may be identified:

(a) Teachers make no particular attempt to adapt their teaching of English to the specific requirements of the situation, but teach general English from general English textbooks. This is rarely adopted as a deliberate policy, but reflects the reality of many programmes where attempts on the part of English teachers to adapt their teaching have foundered. Among the reasons for such failure are the negative attitudes which English teachers, usually Arts graduates, sometimes have towards technical material (Hutchinson & Waters 1987, pp.162f.), and the genuine difficulties they may experience if using technical texts, particularly as a parallel technical course progresses. In an English for electronics course, for example, it may be quite feasible initially for English teachers to use texts dealing with general principles of electricity with which they may be familiar from their own secondary

education, but as the material becomes more advanced they may find themselves no longer able to cope with the material, and may increasingly refer learners' questions to the subject teachers. This may lead to a lowering of status in the eyes of the learners (*ibid.*, p.164).

Another reason teachers of English sometimes abandon attempts to tailor their teaching to the demands of the science course may be failure to establish good working relationships with subject teachers (Hutchinson & Waters 1987, p.164). This relationship is often marred by negative attitudes on both sides. English teachers' lack of enthusiasm for technical material, combined with feelings of inferiority arising out of their usual lower status position (*ibid.*), may lead to a reluctance to become involved in working out a co-operative approach with the subject teacher. Subject teachers may feel puzzled about the role of the English teacher, and concerned that in dealing with technical material the English teacher, unqualified in the technical subject, does not upset the technical programme through errors, pre-emptive teaching, etc.

An examination of the data generated from the electronics English corpus in this study shows that lexical or syntactic features typically emphasised in general English courses may not be particularly useful to learners dealing with special English material. For example, some syntactic features popular in general textbooks showed a very low frequency of occurrence in the corpus; the present continuous occurred 12 times, the past simple 6 times and hypothetical conditions only 4 times. For ESP learners it hardly seems worthwhile spending much classroom time on these points. In addition, the lexical emphasis, if any, in these general textbooks often seems to be on the names of everyday objects almost never encountered in technical material, as an examination of the data will show. For example, there is a one-page word list entitled "Vocabulary" given at the back of *Building Strategies* (Abbs & Freebairn 1984), a popular general course book for pre-intermediate learners, which is entirely a list of everyday nouns, e.g.: bookshop, museum, cupboard, lettuce, jeans, socks, etc. The selection of these items seems to be based more on their familiarity (Richards 1974, p.77) or availability (*ibid.*, p.75, see also section 2) than their frequency of use in English generally. In addition, the electronics English data shows that even common general words may be used in special ways in technical material. For example, the word *since* is found 22 times in the material as a discourse marker meaning because, and not at all to relate a point of time in the past to the present, the way in which it is typically presented on general courses.

(b) A modification of the above approach is seen where the English teacher restricts himself to teaching general English and using general material, but gives priority to syntactic forms which are more likely to occur in scientific writing: the passive, nominal compounds (i.e. noun/noun constructions: *a car door, a boat engine*, etc.), relative clauses. However, the chief distinguishing features of special languages seem to be lexical rather than syntactic (see Introduction above).

(c) Teachers use "common core" English for Science coursebooks, such as *First Course in Technical English* (Beardwood et al. 1978) or *Elementary Technical English* (Webber & Seath 1983). This type of coursebook attempts to meet the needs of ESP learners in general by devoting each unit of the book to a different scientific area, focusing on syntactic structures such as the passive, nominal compounds, etc., which have been found to have a high frequency of occurrence in scientific and technical writing. The problem with textbooks of this type is again lexical; a text on industrial chemistry, for example, such as occurs in one unit of *First Course in Technical English, Book 2* (Beardwood et al. 1979), will contain technical vocabulary items specific to chemistry. This creates an unnecessary extra learning burden for learners who are following a course in a technical area other than chemistry (this point is discussed further in section 4).

(d) Teachers may combine the above approaches with using authentic subject-specific material taken from the subject textbook, or handouts used by the subject teacher, in reading comprehension classes. The chief difficulty here is the technical vocabulary contained in such texts, which may make the material inaccessible to the English teacher (see above). Hutchinson & Waters note that the English teacher is at a serious disadvantage in "trying to explain specific vocabulary without the visual aids and expertise that the technical college can produce" (1981, pp.56f.).

(e) Occasionally, an ESP teacher may develop a genuine interest and expertise in the subject. This may be the happiest of all solutions, the only danger being that in his/her new-found enthusiasm for the subject the teacher may neglect the teaching of language-specific skills, and become in fact a subject teacher.

(f) Finally, sometimes an ESP teacher may have a technical background, and be therefore both motivated and qualified to teach the subject along with teaching English. This type of individual is a rarity, however, and likely to remain so, for the simple reason that subject

teachers are almost always paid more than their English teaching colleagues.

The kernel of the question is to what extent and in what specific ways the ESP teacher should adapt his/her teaching to suit the requirements of the subject course. Should it be part of the ESP teacher's job to learn as much as possible about the subject, and if so, how should this knowledge be put to use? Hutchinson & Waters suggest that the ESP teacher should take an active interest in the subject, but that this should not mean that he/she becomes a teacher of the subject matter, but rather an interested student of the subject matter. They feel that it is perhaps more important that the ESP teacher maintains a positive attitude to the subject, as a negative attitude on his/her part may affect learners' motivation to learn the subject (Hutchinson & Waters 1987, p.163). They then go on to ask a pertinent question: what about the not unusual situation where the ESP teacher is faced with a class of learners with different subject specialisms? (*ibid*., p.165). Clearly, in this case a subject-specific approach is not feasible. Kennedy & Bolitho mention the same difficulty and suggest a common core approach in this situation: "Texts of a semi-technical nature may be chosen, perhaps drawing on topics (such as pollution, population, conservation, etc.) ... which would provide practice in a set of skills, structures, functions and semi-technical vocabulary which the students will meet in their specialist studies" (1984, p.50).

In summary, then, it seems clear that there is a need to elaborate a common core approach to ESP courses to help clarify the role of the ESP teacher, in particular to help the ESP teacher who may be unable or unwilling to become involved with the subject, and also to suggest a way forward for the teacher who is faced with a class of learners from various subject specialisms. We have also seen that a common core approach based on the syntactic features of subject specialisms is in fact often used by ESP teachers and is the organizing principle behind many existing ESP textbooks. Less common, and potentially more valuable for learners, are approaches based on the lexical "common core" of special languages.

2 Semi-technical vocabulary

A number of writers in ESP have mentioned the notion of semi-technical vocabulary (sometimes called "sub-technical vocabulary"), often listing half a dozen intuitive examples of such words (e.g. Kennedy & Bolitho list the items *reflection, intense, accumulate,*

tendency, *isolate* and *dense*; 1984, p.58), but few have attempted to elaborate a comprehensive list of such items based on empirical analysis, or made a serious attempt to define precisely what is meant by the term "semi-technical vocabulary". These are objectives of the present study.

C.L. Barber may have been one of the first to investigate the notion of a general vocabulary of science, defining it as "high frequency words not contained in West's *General Service List of English Words* (*GSL*) (1953)" (Sager et al. 1980, p.242) in a frequency study of 23,400 words from three texts (electronic engineering, biochemistry and instrumental optics). His general scientific vocabulary included:

nouns:	*equation, method, phenomenon, principle, process, series*
qualifiers:	*accurate, complete(ly), considerable(y), constant, definite, positive, similar, sufficient(ly)*
verbs:	*alter, approach, assume, consist, indicate, obtain, occur, require, vary*

(Sager, Dungworth & McDonald 1980, p.242)

Swales notes that Barber, "Like practically everybody else involved in English for Science and Technology, ... holds the view that teaching specialized technical terms falls neither within the responsibility nor the competence of the English teacher" (1985, p.17). Swales goes on to quote Barber himself: "What the English teacher can usually hope to do is to teach a vocabulary which is generally useful to students of science and technology - words that occur frequently in scientific and technical literature of different types" (*Ibid.*, p.18).

Kennedy & Bolitho suggest that the technical vocabulary of subject specialisms arises in context in the specialist classes and is not the language teacher's responsibility, while "sub-technical" vocabulary, or "words which are not specific to a subject speciality but which occur regularly in scientific and technical texts" and with which learners frequently find difficulty, should be given high priority in the language programme (1984, pp.57f.). Higgins also points out that "it is not the job of the English teacher to teach technical vocabulary; it consumes too much time, and he will probably not do it well" (1967, p.32). Instead he suggests that the English teacher concentrate on "frame" words, a list of which he drew up in association with subject teachers, and defined as "words which were causing difficulty, words which, although not technical terms, are frequently used in technical writings". He notes that the scientist cannot teach these words properly (*Ibid.*,

p.32). This is the list he presents, which does not seem to have been based on a frequency count (he does not mention which areas of science these items were taken from):

contain	*increase (n. & v.)*	*separate*
include, including	*decrease (n. & v.)*	*combine*
consist of	*obtain*	*boil/heat*
record (n. & v.)	*determine*	*fill*
consume	*react*	*assemble*
materials	*chemicals*	*proportion*
quality	*figures*	*theory*
average	*exceptional*	*theoretical*
similar	*identical*	*pure*
accurate	*inaccurate*	*impure*
exact	*approximate*	

Comparing Higgins' list with Barber's above, we find only two items which occur on both lists: *obtain* and *accurate*.

Robinson suggests an important role for semi-technical vocabulary: "Coursebooks ... perhaps do not need to concentrate on the very specialized vocabulary items as students will get these from other sources. Rather it is the sub-technical level which is often difficult". She notes that Phillips et al. found in a study of four agriculture textbooks that only about 15% of verbs were specifically associated with agriculture, whereas 60-70% were semi-technical, i.e. generally applicable to science (Phillips et al. in Robinson 1980, p.71).

Wallace makes the point that an ESP learner who is already knowledgeable in his subject does not have to grapple with technical concepts, but rather "learn ... the label to attach to a concept that he is already familiar with". He continues: "The serious problem ... is probably not technical language as such, but the language framework in which the technical expressions are placed ... sub-technical words and expressions typical of academic discourse (that is, words such as *ratio, approximate, hence*, etc.) which the subject specialist may assume that the student should already know" (Wallace 1982, pp. 171.).

Inman, in a study of over 100,000 words of scientific and technical prose, distinguished between technical and sub-technical forms, a distinction which was "essentially intuitively determined and was based on the definition of subtechnical vocabulary as context-independent words which occur with high frequency across disciplines". She found that 9% of occurrences were function words, 21% technical vocabu-

la and 70% sub-technical vocabulary. She concludes that it is "sub-technical vocabulary which should be focused on in teaching scientific and technical English. Technical vocabulary ... is best left to presentation through the discipline itself" (Inman 1978, pp.246ff.).

In discussions of semi-technical vocabulary it is often suggested that items of this type present particular difficulty for learners (e.g. Robinson, Higgins, Kennedy & Bolitho above). Richards presents some interesting psycholinguistic evidence to support this view, quoting Michea (1953) who noted three facts concerning vocabulary:

(i) After being read a text, students find it easier to recall the concrete vocabulary in it than the structural words.

(ii) In certain aphasic conditions, words disappear in a definite order: first proper names, then the names of things, then abstract nouns, then verbs, and lastly grammatical words.

(iii) When asked to list the vocabulary associated with a certain topic, concrete nouns come to mind more easily.

Richards notes Michea's conclusion that these three factors illustrate the availability of concrete vocabulary (Richards 1974, pp.75f.). Since semi-technical items seem to be context-less and non-concrete (see below, and Appendices A & D), and therefore less available in Michea's terms, we might conclude that they may indeed present greater difficulty for learners than technical vocabulary, which often refers to concrete items of equipment, etc., suggesting an important role for semi-technical vocabulary in ESP courses. Carter also notes that "Concrete words are generally learned first and are generally easier to retain and recall. Abstract words may be more difficult" (1987, p.152).

As we have seen, a number of ESP writers have expressed interest in the idea of semi-technical vocabulary. However, some of the definitions suggested by these writers are less than rigorous, and in many cases only a handful of examples are given, based on intuition, before the writers go on to discuss something else. How are we to decide, for example, just what items are "typical of academic discourse" (Wallace 1982, p.18), once we have come to the end of Wallace's list of three words? Barber made an early attempt to elaborate a list of semi-technical items based on empirical research, but the study covered only three different scientific disciplines with a corpus of about 6,000 words in the case of each, a very small sample from which to make generalizations about all sciences. Inman carried out a study of 114,460 words taken from ten disciplines (biology, mathematics, physics, chemistry and chemical engineering, geology, mining engi-

neering, electrical engineering, civil engineering, mechanical engineering, metallurgical engineering), a considerable improvement on Barber (Inman 1978, p.243). Unfortunately, the complete results of her research do not seem to be available in published form. In her article in Todd Trimble 1978 she lists only the ten most frequently occurring sub-technical words: *high, system, result, process, function, form, temperature, large, solution, structure* (Inman 1978, p.247).

Johansson has carried out an empirical study of the vocabulary of learned and scientific English, using the data generated from the Brown University Corpus study (Kucera & Francis 1967). The Brown Corpus consisted of over a million words, of which 160,000 fell into the category of learned and scientific writing. Johansson elaborated a list of items "characteristic of SE" (Special English) based on a "differential ratio" which related the frequency of occurrence of items in Special English to their frequency in the general corpus (Johansson 1975, p.6). The differential ratio could be used to define semi-technical terms, though there are some difficulties with this. (See below for a further discussion of Johansson's work).

Yang has recently made an important contribution to the discussion of semi-technical vocabulary, in an analysis of the lexical characteristics of a relatively large corpus of 300,000 words of scientific text. Yang noted that items with a high distribution (i.e. number of text samples in which the item is found) but relatively low frequency seemed to represent notions general to all or most of the subject areas, and called these "sub-technical" words. (Yang 1986, p.98). (This approach is also discussed in more detail below).

It seems that the concept of semi-technical vocabulary has perhaps not received quite as much attention as it deserves. This study, then, is in part an attempt to fill this gap; specifically, to formulate a more precise definition of semi-technical vocabulary, to propose an extensive list of such items as a reference tool, and to explore what role semi-technical vocabulary might have in materials design for ESP courses.

The definition of semi-technical vocabulary which I propose, and which I justify in detail below, is as follows: Formal, context-independent words with a high frequency and/or wide range of occurrence across scientific disciplines, not usually found in basic general English courses; words with high frequency across scientific disciplines

A review of the various definitions of semi-technical vocabulary given above shows that most identify frequency of occurrence across different scientific areas as a basic criterion. As we have seen, Johansson

used frequency of occurrence in different disciplines in elaborating a list of items characteristic of Special English. He took the 1,000 most frequent words on the list from Category J in the Brown Corpus, which was a subdivision of the main corpus consisting of 160,000 words in the area of learned and scientific writing. He then compared this list with the 1,000 most frequent items on the list generated from the entire corpus. Over 300 items from the Category J list did not occur at all among the 1,000 most frequent items generally, presumably marking them as characteristic of Special English, though Johansson does not comment on the significance of this. These words include such items as *defined*, *contain*, *corresponding*, *independent*, which intuitively seem characteristic of science generally, but also some subject-specific items: *hypothalamic*, *communism*, *philosophy*, *pulmonary* (Johansson 1975, pp.43-8).

Johansson then proposes a differential ratio to establish which items are characteristic of Special English. This is determined by dividing the number of occurrences of an item in the Category J material by the number of occurrences of the item in the corpus as a whole, and multiplying by a hundred, giving a percentage figure. For example, the item *sample* was found 48 times in the Category J material, and 57 times altogether, yielding a differential ratio of 84. Johansson continues: "As the material in Category J (80 samples, 162,117 words) accounts for approximately 16 per cent of the whole corpus, a differential ratio of around 16 for an item means that it has approximately the same frequency in the SE material as in the whole corpus. A considerably higher or lower figure indicates that a word is characteristic or non-characteristic of SE, respectively" (*Ibid.*, p.6). The 1,000 most frequent items on the Category J list are then presented, arranged in groups according to differential ratios: words of ratio 0-11, 12-22, 23-33, and so on. Words with a low differential ratio are part of the general lexicon, e.g. *he*, *I*, *what*, *who*, have a ratio of 0-11. Johansson notes that the group of words of ratio 23-33 and subsequent groups "represent, at an increasing rate, words typical of SE" (1975 p.16). The majority of the words on the lists of ratio 23-33, 34-44, 45-55 and 56-66 (*Ibid.*, pp.15-20) seem to be part of the general vocabulary of science on intuitive criteria: items such as *information*, *process*, *analysis*, *surface*, *description*, *section*, etc. As the differential ratio increases, however, the word lists become shorter, and most of the items are now subject-specific. Here is Johansson's final list, items with a differential ratio of 89-100:

bronchial	*detergent*	*polynomial*
hypothalamic	*dictionary*	*questionnaire*
optimal	*emission*	*sampling*
pulmonary	*fiber*	*shear*
thyroid	*foam*	*staining*
anode	*foams*	*substrate*
binomial	*gyro*	*substances*
carbon	*index*	*tangent*
chlorine	*onset*	*urethane*
coating	*oxidation*	*variables*

(Johansson 1975, p.21)

So it seems, as Johansson acknowledges, that "terms peculiar to specific sciences appear with a very high differential ratio" (1975, p.21). We could, then, use Johansson's differential ratio as an objective means of identifying semi-technical vocabulary, where the ratio of an item falls in the middle range, between 23 and 66. However, if we inspect these lists in Johansson (*ibid.*, pp.15-20), we can see that even here quite a high proportion of items are subject-specific, e.g. *social, education, funds, market, government, school, teacher, poetry, justice, religion*. This may be due to the lack of any information in Johansson's study concerning the range of items, or the number of different text samples in which they occur (see below). Johansson admits that this is a difficulty (*ibid.*, p.3). However, we should note that he was not primarily concerned with identifying the general vocabulary of science, nor were his aims pedagogical.

Frequency of occurrence across scientific disciplines must remain a basic test of which items are characteristic of science generally. Johansson only concerned himself with the 1000 most frequent items on the Category J Brown Corpus list in his discussion of the differential ratio. The data from the LOB Corpus study (Hofland & Johansson 1982) also includes a list of the most frequent items in a number of different sciences, with the advantage that a measure of range is given for every graphic word listed. It was therefore decided to use this data rather than Johansson's in compiling a list of semi-technical vocabulary, as range was felt to be of particular importance in elaborating a definition of the term.

The LOB study came later than the Brown Corpus, which was an American English corpus, and was intended as the British English equivalent. It was organised along similar lines, with the aim of making the data from the two studies comparable (Hofland & Johansson 1982, p.1). Thus it also has a Category J, learned and scientific writing, which

also contains 80 text samples of 2,000 words each, giving a corpus of 160,000 words. Text samples were taken from the following subject areas: natural sciences, medicine, mathematics, psychology, sociology, demography, linguistics, education, politics, economics, law, philosophy, history, literary criticism, art, music, technology and engineering. The frequency list for category J in the LOB Corpus should certainly give us words with high frequency across scientific disciplines. However, as with Johansson's data, some frequent items are obviously context-specific: *school*, for example, has a frequency of 123.

2.1 Words with a wide range

Widdowson notes that in West's *General Service List of English Words* (1953) "it was found that certain lexical items of high aggregate frequency also occurred across a wide range of texts" (1983, p.92). He uses the term "procedural vocabulary" for this type of item, and goes on to illustrate its role in defining specialised terminology: "... words of wide ... range are especially useful for ... defining terms which relate to particular frames of reference ... such words function as a procedural vocabulary for establishing the terms which characterise different schemata" (1983 p.93). Widdowson gives the following entries from the *Oxford Advanced Learner's Dictionary of Current English* to illustrate his point:

pinnate: (*bot.*) (of a leaf) formed of small leaves on opposite sides of a stem

pipette: slender tube for transferring small quantities of liquid, esp. in chemistry

Widdowson notes that the terms *pinnate* and *pipette*, specific to botany and chemistry, are defined by the use of common core or procedural words (1983, p.93); semi-technical vocabulary, in our terms. Richards also notes the importance in pedagogical word lists of items which have "definition value", i.e. "words which help define other words" (1974, p.74).

The word *school*, in spite of its high frequency in the LOB Corpus, is intuitively not a word likely to be encountered in all scientific disciplines. But its low range, 13 out of 80 text samples, is an objective indicator of its restriction to a limited number of subject areas. *Blood*, with a frequency of 83, is another example. Again, it has a low range of 5. Generally speaking, high frequency items which are clearly not part of the common core vocabulary of science but are characteristic of particular disciplines seem to reflect this in their low range: *oxygen*

(63/6), *coal* (46/5), *sodium* (60/5), *trade* (63/11). Frequency alone, therefore, is an insufficient indicator of words prominent in science generally. To be significant in this sense an item should also have a high range.

The maximum range in the LOB Corpus category J is 80, therefore a high range figure means somewhere near the top of a scale from 1 to 80. But of course the range figure is limited by the frequency figure; clearly, an item with a frequency of 10 can only have a maximum range of 10. Some items of the latter type, in spite of their low frequency overall, seem to have significance in a list of semi-technical vocabulary. The word *otherwise*, for example, occurs only 15 times but in 15 text samples; i.e. it occurs with "maximum" range. Where the range is maximum, or near maximum, this may be an indicator of items which are significant in the common core vocabulary of science, even though their absolute measures of range and frequency are low. If we add to this objective measure an intuitive evaluation, *otherwise* seems like a word generalizable to all areas of science, and therefore part of semi-technical vocabulary. Range figures which are relatively high, then, may also be significant. Other examples might be *throughout*, with a low frequency of 24 but a near maximum range of 20, *obvious* (23/19), *noted* (22/19), *whereas* (24/19), *established* (23/19), *related* (26/22) (see Appendix A for further examples).

Yang (1986, p.98) notes a similar characteristic of "sub-technical" words, observing that they have high distribution but relatively low frequency. Although Yang's corpus of 300,000 words is larger than the LOB corpus, it is divided into only 10 text samples, thus giving a less subtle estimate of distribution. On the other hand, Yang's data gives individual frequency figures in each text category, thus giving an insight into "evenness" of distribution. The list of sub-technical items Yang proposes (*ibid.*, p.95) is shown below. Frequency figures in each text category are given (the row of figures across the top refers to the ten text samples):

	1	2	3	4	5	6	7	8	9	10
absolute	3	14	7		34	7	14	5	17	
accuracy	5	28	8	6	11	3	17		18	
act	5	65	15	11		8		18	33	5
basic	18	6	13	8	18	15	3	53	25	
basis	2	7	18	4	19	1	17	7	9	
bearing	1	2		1	1	3	5	9	1	1

	1	2	3	4	5	6	7	8	9	10
calculate	25	35	51	26	91	34	48	15	47	
carry		18	4	11		5	1	19	5	3
conclusion	2	1	8	5	11	6	10	4	5	3
decrease	13	7	10	13	31	15	12	39	32	
define	16	6	11	6	16	31	5	24	12	
definition	7	25	15	5	61	15	17	13	26	1
effect	63	144	103	43	42	90	28	43	24	1
effective	27	21	13	26	3	8	1	13	1	
electrical	6	58	27	82	26	1	22	11	14	
face	2	6	4	1	2	2	4	2	9	34
fact	12	31	59	6	22	68	51	50	43	14
factor	50	7	46	46	29	27	4	22	20	
feature	24	3	5	8	4	8	2	8	8	1
give	145	71	120	53	42	23	101	21	95	37
increase	19	43	18	11	89	21	21	50	53	1
introduce	2	7	7	2	6	16	3	1	6	3
know	26	9	12		26	10	11	4	12	181
possible	48	42	45	14	22	35	42	52	60	14
problem	21	4	117	20	54	81	9	7	64	2
result	47	92	135	29	70	49	79	19	120	
total	17	98	85	30	39	76	27	30	85	2

2.2 Context-independent (non-technical) words

Widdowson defines common core or procedural vocabulary as words which have a high frequency and wide range (see above), and notes that "these 'common core' items are not schematically bound" (1983 p.92). As noted above, in most cases context-specific items identify themselves by their low range; thus *investment*, clearly a term from economics, occurs 32 times but in only 5 out of 80 text samples. However, in some cases items which by intuitive criteria seem bound up with a particular context have a not insignificant range: *war*, for example, has a range of 21. Other items have a low range absolutely and relatively and yet intuitively seem generalisable: *particularly*, for example, has a frequency of 53 and a range of 9.

Yang notes that his sub-technical words represent notions general to all or most of the subject areas, as most of them seem to do; however, the word *electrical* seems out of place in such a list. Yang's corpus seems somewhat biased in the direction of physics, which

might account for this, together with his concern to elaborate objective statistical definitions only. But these examples illustrate the need to apply subjective as well as objective selection criteria to arrive at a list meaningful in pedagogical terms, or perhaps even in theoretical terms. Richards, in an article on frequency lists, has also made this point: "... subjective and objective measures of vocabulary selection need to be combined if the validity and utility of such lists are to be improved" (Carter & McCarthy 1988, p.10). If we are to isolate context-independent words then we must eliminate a few words with a relatively high range which we intuitively determine to be characteristic of a particular subject area, as well as those with a low range, while a few low range items may be included as semi-technical items on subjective criteria. However, objective measures of frequency and range (including relative range) supply most of the items on our semi-technical list; the number of items to which we must apply subjective criteria is small.

2.3 Words not found in general courses

If we used the definition of high-frequency words alone, we would then include the frequently-occurring grammatical words which head the category J list: *the*, *and*, *of*, *in*, etc. Barber eliminated basic or general words by not including words found in West's *General Service List of English Words* (1953) in his list of the general vocabulary of science. Our elimination of general words will be based on an intuitive evaluation of what words would be likely to occur in general or basic English courses to a lower intermediate level. In addition to the very frequent grammatical words mentioned above, many other general words occur among the first 100 items on the LOB list: *can*, *will*, *would*, *should*, *which*, *where*, *what*, for example.

A more objective way of defining general words might be to apply some of Carter's tests for core vocabulary (1987, pp.35-42), as follows:

1. *Syntactic substitution:* In this test a word is defined as core if it can substitute for others, but not vice-versa, thus *eat* can substitute for *gobble*, *dine*, *devour*, etc. (Carter 1987, p.35). An example from the electronics material would be *make*, which can substitute for *produce*, *generate*, *construct*, *build*.
2. *Collocability:* Here an item is said to be core if it contracts a wide range of collocations and fixed expressions (Carter 1987, p.36). In this study grammatical words like *in* and *as* seemed to have the highest number of different collocations (see Farrell 1989, Appendix F).
3. *Associationism:* This test identifies core words where they fall in the

mid range of assessment scales of formality/informality, strength/weakness, and positive/negative meaning, as rated by informants (Carter 1987, p.40). Certainly many of the words on our general list would seem to be neutral in terms of formality, and contrast in those terms with semi-technical equivalents (see below), e.g. *easy* v. *simple*, *keep* v. *maintain*.

4. *Frequency:* Carter also notes that frequency of occurrence is a central criterion of core vocabulary (1987, p.43). Many of our general words are among the most frequent in the material.

While two of Carter's tests could be successfully applied to semi-technical words, i.e. frequency and a test for "neutral field of discourse", which seems to identify context-independent items (1987, p.41), semi-technical words would fail on many of his other tests. Indeed, a further element of the definition of semi-technical vocabulary seems to be that it is non-core vocabulary; elsewhere Carter refers to the notion of core vocabulary as a "basic core vocabulary for initial language learning purposes" (1987, pp.33f.), a definition which clearly cannot be applied to semi-technical words.

2.4 Formal words

Most writers interested in semi-technical vocabulary seem to list relatively formal items, though this does not usually form part of the definition; in compiling a list of semi-technical vocabulary *obtain* will be preferred to *get*, *produce* to *make*, *occur* to *happen*, etc., thus failing Carter's test of "associationism" above. Mackay & Mountford note that the vocabulary of ESP often consists of words which are "more formal equivalents to words and phrases of the "common core" of English, e.g. *... residue, derived from, composed of*" (1978a, p.145). Herbert also drew attention to the formality of semi-technical vocabulary: "... verbs, adjectives and adverbs that are not specifically scientific, but which belong to the phraseology of science. These are usually formal, dignified and foreign-sounding words, like *extend* and *propagate* and *obviate* and *negligible* ..." (Herbert in Swales 1985, p.18).

If we then apply all the above selection criteria to the LOB Corpus category J list (Hofland & Johansson 1982), eliminating general, context-specific and low frequency/range items (i.e. all items with a frequency of less than 20), we arrive at the list of items which may be found in Appendix A, and which is proposed as a list of the general vocabulary of science, or semi-technical vocabulary. The ten most frequent items on this list are: *per cent*, *however*, *form*, *point*, *results*, *field*, *area*, *possible*, *even*, *general*. Two of these correspond with

Inman's list of the ten most frequent sub-technical items: *form* and *results*. Of the 26 items on Barber's list of the general vocabulary of science, 16 are found to occur on our list: *method, principle, process, series, complete, completely, considerable, constant, similar, sufficient, sufficiently, approach, indicate, obtain, occur, require*. Of Yang's list of 27 sub-technical items (see pp.15f. above), 14 occur on the list derived from the LOB corpus, and are all significant items in terms of frequency and range: *accuracy, act, basis, definition, effect, effective, fact, factor, increase, know, possible, problem, result, total*. Using only intuitive criteria, Higgins included in his list of 34 items 12 which occur on the LOB-derived list: *including, materials, average, similar, increase, obtain, determine, figures, separate, proportion, theory, theoretical*.

3 A lexical study of the English of electronics

3.1 Vocabulary in ESP

We have mentioned above the long-standing neglect of vocabulary in language teaching. Candlin feels that this is due, in part, to "the overwhelming concentration in linguistic theory ... on issues of syntactic structure". He goes on to point out, however, that "the study of vocabulary is at the heart of language teaching and learning, in terms of the organization of syllabuses, the evaluation of learner performance, the provision of learning resources, and, most obviously, because it is how most learners see language and its learning difficulty" (1988, p.vii). Undoubtedly syllabuses in EFL coursebooks have tended to look to either syntactic categories or language functions for an organizing principle. Renouf, in a survey of nine major EFL courses found that the vocabulary was regarded "merely as the means of exemplifying other features of the language ... it is not ... organised in and for itself ...." She notes that while word lists are often included in coursebooks, it is not clear what this signifies: "Many words which occur several times in the body of the book are not acknowledged at all, whilst official teaching words sometimes receive very little reinforcement, with some occurring only twice in the entire volume" (Sinclair & Renouf 1988, p.142). Kennedy & Bolitho stress the particular importance of vocabulary in ESP courses: "... it is dangerous - perhaps more so in ESP than in general English - to assume that vocabulary will take care of itself ..." (1984, p.56). However, Hutchinson & Waters (1987), in an otherwise excellent book on ESP, devote only a quarter of a page to the subject of vocabulary.

Swales notes that in LSP "the lexical level ... has been neglected" (1983, p.22). He goes on to evaluate some relatively recent ESP coursebooks in terms of their treatment of vocabulary. He is scathing about the preoccupation some ESP course writers have had with reference items: "If we examine the *Focus* series (Allen & Widdowson 1974), we can see that comprehension work on individual lexical items is largely devoted to ... items such as *it, this, them*" (1983, p.22). Swales queries the extent to which such reference items present a real language problem for learners, and suggests that this type of material "was included at the expense of neglecting to ensure that the learners had a level of vocabulary sufficiently high for them to undertake ... relatively advanced reading tasks" (1983, p.22). He notes the same neglect in *Reading and Thinking in English* (Widdowson 1979) and *Skills for Learning* (Morais 1983): "... long experience indicates that it is counter-productive to attempt to develop advanced reading skills on texts that contain even a single-figure percentage of words that are not known by the readers" (1983, p.23).

If we accept the need for a lexically principled approach in syllabus design, we must then address the issue of how to go about selecting items for inclusion in a word list which will be useful pedagogically. Intuitive criteria have been used, for example Richards's "Basic English" list of 850 words (Carter 1987, pp.221.). Carter, as we have seen, has proposed the notion of core vocabulary. One of Carter's criteria for defining core vocabulary is frequency of occurrence (see p.18 above), a basic criterion often used in drawing up pedagogical word lists. Sinclair & Renouf suggest that "It seems reasonable ... to propose that, for any learner of English, the main focus of study should be on the commonest word forms in the language ..." (1988, p.148). In addition they note the importance of the central patterns of usage of these words, and the combinations which they typically form (1988, p.148). Sinclair has been involved in the COBUILD project at Birmingham University, a large-scale frequency count of 20 million words of spoken and written text. The data from this study has been used in compiling a dictionary and in forming a lexical syllabus for a series of EFL coursebooks (Willis & Willis 1988). In reviewing the COBUILD project Carter notes that "knowing the most frequent words and the predominant patterns into which they enter is essential for language learners and their teachers" and "introspective or intuitive knowledge on the part of native speakers may not be sufficient" (Carter 1989, p.33).

Sinclair & Renouf note that the results of frequency studies sometimes seem counter-intuitive, for example the finding from the CO-

BUILD study that the first and second most frequent uses of *see* occur in the phrases *you see* (an indication of interactive concern in spoken discourse) and *I see* (a response meaning "I understand"), rather than seeing through one's eyes (Sinclair & Renouf 1988, pp. 151f.). This underlines the weakness of using intuitive selectional criteria. Similarly, in the present study of electronics English *now* was found to occur 33 times as a discourse marker, meaning in fact "then", "after that", and only once in the sense which might be expected to be central or basic for this word: "today", "at present".

Sinclair & Renouf point out that most attempts to create lexical inventories for syllabuses have been concerned with general purpose English programmes, while the needs of a specific group of learners are usually easier to identify (1988, p.148). One of the difficulties they acknowledge is the fact that the most frequent words in the Birmingham Corpus are function words and low in information content: "... it would be difficult to construct a motivating course based entirely on the 200 words listed above" (Sinclair & Renouf 1988, p.150). By contrast, a glance at the first 200 items in the list from the present study of Electronics English (see Appendix C) shows a rich variety of items which could be used in materials.

Discussions of needs analysis in ESP seldom give much attention to lexis; in Hutchinson & Waters' chapter on the subject the "lexical features" of target situations are mentioned in passing, and that is all (1987, pp.53-64). But lexical research of the COBUILD type has even greater potential for ESP than general English courses, as it is much easier to identify recurrent lexical features of restricted registers; this type of research may be seen as a kind of lexical needs analysis. Praninskas notes that "For as long as I have been teaching English, language students have been badgering me for lists of important words" (1972, p.7). Accordingly, Praninskas decided to prepare a list of words of high frequency in the textbooks students were required to read in the first year of study at the American University of Beirut. This list was used as data for teachers and materials writers planning vocabulary lessons, and in developing a vocabulary course. Praninskas concludes that "The vocabulary course ... has been taught for the past two years with most gratifying results. Students enthusiastically report the relevance of the words in their lessons ..." (1972, pp.7-9). In a similar way, it is hoped that the following study of the English of electronics can make a contribution to subject-specific ESP course design.

3.2 The corpus

The corpus consists of ten 2,000-word samples of running text in the area of electronics (including basic electricity), totalling 20,000 words in all, each text sample being taken from a different textbook. The following criteria were used in the selection of text samples:

(a) Relevance: From the author's experience of participating as an ESP teacher on a training programme for electronics technicians, texts were selected which seemed typical of the material likely to be encountered by learners on such programmes.

(b) Style: An informal inspection of a number of textbooks used in secondary schools and universities showed a variation in writing style, with consequent variation in the lexicon, from the almost conversational style of *Physics is Fun* (Jardine 1972: ref. in Appendix E) to the formal scientific prose of *Essentials of Higher Physics* (Webster 1963: ref. in Appendix E). Accordingly, an attempt was made to include text samples representative of different writing styles.

(c) Topic variation: Again, an informal inspection of available material showed that a particular sub-field of study within the broad area of electronics would yield items with a high frequency of occurrence which might not be met at all in texts dealing with other topics. This perception was confirmed by the data: for example, the item *transistor* showed a relatively high frequency of 56 occurrences, but was found in only 2 of the text samples. This underlines the importance of a measurement of range even within a single register in deciding what lexical items should be given priority in a subject-specific programme. To avoid skewing the results in favour of any particular sub-field in electronics (e.g. the oscilloscope, radio) an attempt was made to include text samples from as many different areas as possible. The text samples which form the corpus are listed in Appendix E.

3.3 Collection and presentation of the data

In any study of lexical frequency, the first issue to be settled is what exactly we mean by an occurrence of a word. In the American Heritage frequency study (Carroll et al. 1971) the following definition of a word was adopted: "A word is defined as a string of graphic characters bounded left and right by space" (*ibid.*, p.xiii). The authors acknowledge the limitations of such a definition: "This definition is insensitive to differences in meaning and function; it treats all words spelled the

same way as the same word" (*ibid.*, p.vi). If this definition had been used for the present study the separate occurrences of the verbs *conduct* (electricity) and *conduct* (an experiment) would have been listed as an occurrence of the same item, as would the occurrences of *about* as a preposition and as an adverb meaning "approximately". This is clearly unsatisfactory; such semantic information is essential if the resultant word list is to be useful pedagogically (see Farrell 1989, Appendix E). The LOB Corpus study used a similar definition: "A graphic word is defined as a sequence of alphanumeric characters surrounded by spaces" (Hofland & Johansson 1982, p.7). Again the authors acknowledge its deficiencies: "Our rank list shares the weaknesses of other similar listings. As it is based on graphic words, it does not specify ... the frequency of the various senses of lexical items" (Hofland & Johansson 1982, p.20).

A second question which such a definition raises is whether it makes sense to separate occurrences of related derivational and inflectional forms; for example, *varying*, *varies*, *vary*, *variation*, *varied*, *variable*, *various*, *variety* would be listed separately in a frequency list based only on the above definition. They would, however, fall together naturally in an alphabetical list. But forms related inflectionally and spelled very differently would not be found together on either a frequency or an alphabetical list; for example the various forms of the verb *to be*, whether as main verbs or auxiliaries: *is*, *be*, *are*, *was*, *being*, *been*.

In addition, such an approach gives us no syntactic information; so, for example, the form *were* as a main verb in the past simple, as an auxiliary in the past simple passive, and as a subjunctive verb in a hypothetical conditic would all have been counted as the same occurrence.

And finally, multi-word items which form a single semantic unit would be broken up and listed separately. This type of item seemed to occur much less frequently in the electronics corpus than in general English, but was found: *come to rest*, for example.

Inman, in her frequency study of scientific and technical prose (see pp.9f. above), linked derivational and inflectional bound morphemes and reduced them to a single base form for the main list, while keeping a record of the bound forms for use in constructing vocabulary and reading comprehension exercises (1978, p.243). It seems clear that from a pedagogical point of view inflectional and derivational forms are best linked, so a similar approach has been adopted in presenting the data from this study: that is, one of the lists (Appendix C) is a

condensed list of base forms or "lemmas" (see below), while the alphabetical list (Appendix B) gives all the graphic words which have been grouped under a particular lemma. Detailed semantic and grammatical information for each graphic word may be found in Farrell 1989, Appendix E.

It seems from the above discussion that not one, but several definitions of the term "word" are needed. What do we mean when we say that the corpus consists of 20,000 "words", for example? This, of course, describes the number of occurrences of any graphic words, including repeated occurrences of the same graphic word; so 2034 occurrences of *the* = 2034 words. Some researchers have used the term "token" here, so the 2034 occurrences of *the* would be 2034 tokens. The term "type" is then used to refer to a string of characters which occurred at any time in the corpus, i.e. *the* is one type in the above example (Barber 1962, p.4). The term "type" would then be interchangeable with the term "graphic word", i.e. a unique string of characters which occurred in the corpus. And if, as we have seen above, it becomes useful for pedagogical purposes to group occurrences of inflected and derived forms, we may need to elaborate a definition of a "basic" form to represent all these forms in a condensed list, or to head an entry containing these items. Sinclair uses the term "lemma" to represent all the inflected forms of the verb (1985, p.84), and this seems to be accepted lexicographic practice; for example, *carry* may represent *carry*, *carries*, *carrying*, *carried*. Here *carry* is an abstract or theoretical construct which may have *carry* as one of its graphic word or word form realizations (Carter 1987, p.8). What is not so clear is whether it is also acceptable to use the verb infinitive to represent derivations which belong to word classes other than verbs (Carter 1987, p.10), for example, to use *emit* to represent *emitter* and *emission* as well as *emit*, *emits*, *emitting* and *emitted*. Much is often made of the distinction between inflections and derivations (e.g. Carter 1987, p.10), but this distinction breaks down when we find forms such as *changing* and *charged* being used as adjectives: *changing voltage*, *applied voltage*. Indeed, present participle and past participle adjectives were relatively frequent in the corpus (see 3.6 below). Sinclair makes a similar observation: "... the classification of *declining* as verbal is a misleading convention; 26 of its occurrences are noun modifiers, closest in syntax to the word class adjective" (1985, p.87). And, of course, nouns and verbs sometimes take the same form, e.g. *supply* or *charge*. So there seems to be a case to be made for extending the term "lemma" to include derived forms. However, the question here is

whether such a theoretical construct is a satisfactory way of representing the data gathered in a frequency study. For example, does it make sense to use *resist* to represent occurrences of *resist*, *resists*, *resistance* and *resistor* when *resist* occurred only once but *resistance* 63 times and *resistor* 47 times? And what about a group like *insulated*, *insulator*, *insulating* and *insulation*, where the verb infinitive did not occur at all? Then there are groups of graphic words which do not have a verb infinitive derivative, for example, *accuracy*, *accurate*, *accurately*. As Sinclair points out, "Lemmatization looks fairly straightforward, but is actually a matter of subjective judgement by the researcher. There are thousands of decisions to be taken" (1985, p.84). From a pedagogical point of view the best solution might be to make a subjective decision in each case based on utility in language teaching terms, which was the approach adopted in Farrell 1989. However, this inevitably resulted in inconsistent decisions: for example, *insulate* was used to represent *insulated* (6 occurrences), *insulator* (5), *insulating* (2) and *insulation* (1), where *insulate* did not itself occur. But *collector* was used to represent *collector* (35) and *collection* (10) where *collect* did not occur, because of the relatively high frequency of *collector*. For the present publication it was decided to adopt a simple objective lemmatization policy which is easy to apply and consistent: adopt the most frequent graphic word as the lemma for each group. This means, for example, using *moving* as the lemma for *moving*, *move*, *motion*, *moves*, *movement* and *moved*. This may seem out of line with lexicographic practice; however, as frequency is a basic criterion by which the lists are drawn up, and is important pedagogically, it would be preferable to maintain a relationship with frequency in specifying the lemma. Sinclair notes that "a case could be made ... that the most frequently-encountered form should be used for the lemma" (1985, p.84).

The frequency and range of each graphic word grouped under a lemma has been added together to give an overall frequency and range figure for the lemma. Lemmas are listed in rank order of frequency and range in Appendix C. The lemma then serves as a reference point on the condensed list from which to proceed. One may not agree with the lemmatization policy, but all the graphic words which occurred may in any case be found in the alphabetical list together with the rank position of the lemma in the condensed list. Then, if more information is required concerning the frequency of individual graphic words, or of different senses of a graphic word, or of its different grammatical functions, or of its occurrence in collocations

or idioms, the detailed list in Farrell 1989 may be consulted (pp.132-202). A page from this list is shown below in illustration:

147.	*all* (det.)		18 - 8 - 14761
148.			18 - 7 - 14779
	form		
	(n.) (= "shape")	6 - 2	
	(v.) (= "make")	5 - 4	
	(n.) (= "type")	3 - 2	
	(n. pl.) (= "type")	1 - 1	
	(n. pl.) (= "shape")	1 - 1	
	forms (v.) (= "makes")	1 - 1	
	formed (pass. v.) (= "made")	1 - 1	
			18 - 7 - 14797
	for example		
	(discourse marker)	9 - 6	
	example		
	(n.)	6 - 3	
	(n. pl.)	3 - 3	
	however		18 - 7 - 14815
	(discourse marker)		
149.			18 - 6 - 14833
	acts (v.)	6 - 1	
	act (n.)	4 - 4	
	action (n.)	4 - 2	
	actual (adj.)	2 - 1	
	acting (pres. cont.		
	rel. del.)	1 - 1	
	actually (adv.)	1 - 1	
	case		18 - 6 - 14851
	(n.) (= "eventuality")	11 - 6	
	(n.) (= "casing")	4 - 2	
	(n. pl.) (= "eventuality")	3 - 1	
			18 - 6 - 14869
	down		

(prep.)	12 - 6
(write) down	2 - 2
downward (adj. attrib.)	3 - 1

As graphic words are listed in these entries in order of frequency (and range), the most frequent graphic word, i.e. the lemma, is found first in each entry. (The lemma is the most frequent graphic word, so in the case of *for example* above, where an idiom is the most frequent, *example* would be the lemma). Where there is grammatical or semantic homonymy, separate figures are given. The first figures are frequency and the second range. Overall frequency and range for the lemma is shown at the top right of each entry, with cumulative frequency in the upper right hand corner. All graphic words in an idiom are listed, e.g. *for example*. The figures in these cases refer to the frequency and range of the idiom as a whole. Collocates which seem interesting are listed to the right or left of a graphic word in italics and bracketed, and have been listed and counted elsewhere, e.g. *(write) down* in the illustration. The relevance of such data for pedagogical purposes is shown by the example of the graphic word form, which has two grammatical functions and three meanings.

There are still unresolved lemmatization issues, however. For example, does *downward* in the illustration above belong under the lemma *down*? Should *overcome* and *overheat* be grouped under the lemma *over*, or do they belong under *come* and *heat*? What about *overlap*? Should *lightning* be subsumed under *light*? There seems to be a clear case for not breaking up fixed collocations and idioms, but which graphic word in the idiom should be related to a lemma? Does *for example* belong with *for* or *example*? It seems difficult to implement any consistent lemmatization policy which will deal satisfactorily with cases of this type. Inevitably some purely subjective lemmatization decisions have remained, based usually on pedagogical utility, for example learner confusion between *lightning* and *lighting* seems possible, and so they may best be dealt with together.

One of the drawbacks of using a computer to analyse a corpus is that only graphic words can be counted, as in the case of the American Heritage study, which then relies on citations which give a printout of the surrounding context for each occurrence of an item to supply semantic, syntactic and collocational information (Carroll et al. 1971, p.vi). A similar approach has been used by Sinclair in the COBUILD project at Birmingham University (Carter 1987, p.141). However, we must still make abstractions from citations to arrive at useful pedagogi-

cal data (see the discussion of *produced*, *make*, *connected*, 3.9 below). Carter notes that, according to Francis, the importance of computers is that they can "count things, ... compare things, ... sort things and find things" (Francis, cit. Carter 1987, p.181), but also points out elsewhere that even in the COBUILD project "there is a heavy reliance on the data generated by the lexicographer rather than on the data a computer might provide" (Carter 1989, p.33). In the case of a relatively small corpus such as the present one of 20,000 words, a manual analysis becomes both feasible and attractive, due to the additional problem of converting the corpus into computer-useable form. Accordingly, a manual count was carried out, recording syntactic information along with each occurrence in most cases, as well as semantic and collocational phenomena where these seemed to deviate from the predictable (see below for elaboration of this point).

In the detailed list (Farrell 1989, pp.132-202) grammatical information was recorded, in most cases only where the possibility of grammatical homonymy existed: thus the word *the*, for example, is not followed by any categorisation, while *charge* is recorded as occurring as a singular noun, a plural noun, an adjective and a verb. Measures of frequency and range are recorded in the case of each category.

West's well-known frequency study (West 1953) gave very detailed semantic information for each entry in the list, noting occurrences of each semantic value of an item as a percentage of the total frequency figure. In this case the meanings of items have been noted only where more than one meaning occurred, for example, a *unit* of measurement and a *unit* meaning a piece of equipment, or where the item was used with only one sense, but that sense seemed unpredictable, for example the use of the word *gate* in a technical sense. This last example illustrates the widespread polysemy found between general and technical items, a phenomenon which may be exploited pedagogically. An intuitive observation was that the occurrence of homonyms seemed lower than in general language; a glance at the entries in this study shows that it was not often necessary to add semantic information, by contrast with West's list where a large number of the entries show occurrences of multiple meanings. Sager et al. note that "... homonymy ... is less in special languages" (1980, p.235), a fact which should encourage the ESP learner and teacher.

Time limitations in the case of this study mean that it was not feasible to carry out a wide-ranging collocational analysis. However, where a particular collocational pattern seemed unusual it was recorded. For example, while most occurrences of *to* were either as a

preposition, defining relationships between objects in physical space (210 occurrences), or as an infinitive marker (223 occurrences), *to* was also found in the following collocations: *to (the nearest cm.)*, *to (within 1mm.)*, *(changes) to (a vapour)*, *to (scale)*, *to and fro*. Where the other items in a collocation pattern are recorded in brackets after the main item, this means that these items have all been recorded as separate occurrences elsewhere in the study (thus the first four collocations above). Where a judgement was made that a string of items constituted a fixed collocation, phrasal verb or idiom, e.g. *go into* (= "investigate"), and/or a separation of the items would result in meaningless entries (for example *fro* in *to and fro*), then these multi-word units were maintained intact in the entry and counted as a single occurrence. Whether syntactic and semantic information pertaining to any occurrence was recorded as an abstraction (e.g. (pass. v.), (= "approximately")) or in the form of a citation, was based on a subjective evaluation of which could supply information useful pedagogically in the most economical manner, without obscuring relevant data. Thus a citation for every one of the 210 occurrences of *to* as a preposition was not felt to be particularly helpful. In addition, a full collocational analysis of *produced*, *make* and *connected* was carried out (see 3.9 below).

The main lemmatized list is broken up into three sections: general words, semi-technical words and technical words (i.e. lemmas; Appendix D). This division is intended as an aid to deciding which lexical items should be given prominence in a vocabulary-based course. General words are defined as words which learners coming to an ESP course will in most cases already have met, as ESP courses are rarely, and probably should not be, taught to complete beginners. They are the common grammatical words *the, be, a, of, to, and, in, it, this, by, as*, etc., together with some common verbs: *use, have, can, make, move, see, give, call, would, pass, need, should, find, open*, etc., and prepositions: *across, through, between, above*, etc. Also common adjectives: *small, long, large, strong, important, difficult*, etc. and concrete nouns: *paper, page, picture, test* (all familiar classroom vocabulary), etc. However, this is not to suggest that these items are uninteresting pedagogically; indeed, some common verbs show a range of meanings, for example *make* (see 3.7 below), and some of the function words are found in a wide range of collocational patterns, for example, *in*, *as* and *at* (see Farrell 1989, Appendix F).

The technical words are defined as words which are best left to the subject teacher, either because they are purely technical terms un-

known in general language, e.g. *anode*, *cathode*, *diode*, *filament*, etc., or because they are words which occur generally but have a precise technical meaning best elucidated by the technical teacher, e.g. *energy*, *force*, *wave*, *signal*, etc. Again, however, the English teacher may wish to deal with some of these items, particularly where the general and technical sense of an item are related (see 3.6 below for an elaboration of this point).

We noted in the analysis of the LOB Corpus data in section 2 that items with a high frequency on the list which were obviously subject-specific could often be identified objectively by their low range. It is interesting to note that in the full list from the electronics English corpus, once the very frequently occurring technical items at the top of the list have been passed (e.g. *current*, *volt*, *circuit*, *resistance*, etc.) a number of the technical lemmas which have a relatively significant frequency show a surprisingly low range by contrast. A glance through Appendix C shows that this goes against the general trend for range to decrease gradually as frequency decreases, and is an indication of these items' restriction to a narrow field of discourse within the broad area of electronics. The following lemmas are examples (range figures follow frequency figures): *transistor* (56/2), *collector* (46/2), *anode* (40/4), *radio* (35/3), *induced* (33/3), *balance* (33/1) *filament* (32/4), *solenoid* (27/1), *C.R.O.* (22/2), *amplitude* (17/3), *silicon* (17/2), *modulate* (16/1), *feedback* (15/2).

3.4 Semi-technical vocabulary

The words which seem to be of principal importance in ESP courses generally are semi-technical words, which form the third list. The definition used in selecting semi-technical items was a modified version of that elaborated in section 2 in relation to science generally: Formal, context-independent words with a high frequency and/or wide range of occurrence across text samples in the area of electronics English, not usually found in basic general English courses. This then gives the third of the three sub-lists in electronics English, which may serve as a reference tool for the English teacher in deciding which items should be given priority in materials design, for electronics English learners in this case (Appendix D).

In the electronics English corpus semi-technical lemmas listed in Appendix D totalled 3,487 tokens, or 17.43% of total occurrences of 20,000 tokens. This figure may not seem significant; however, the extremely high frequency of the top-ranking words should be taken into account. The first nine words on the electronics English main list (*the*,

be, a, of, to, and, in, current, it) before the first semi-technical word, *fig.* at position ten, account for 6,146 tokens, or 30.73% of the whole corpus. In addition, 750 lemmas (representing 2,249 tokens, or 11.23% of all tokens) with a frequency of less than 5 were not listed. A significant number of them were also semi-technical.

What is perhaps more significant is the fact that of the three lists of general, technical and semi-technical lemmas, the semi-technical list is the longest (Appendix D). Out of the total of 508 lemmas occurring more than 5 times, 224 were semi-technical, 143 general and 141 technical. Looked at in this way, semi-technical lemmas were 44% of all lemmas, while technical lemmas were only 27.7%. Using a similar method, Inman in her study (see pp.9f. above) found 70% of "base" forms were semi-technical, 21% technical and 9% function words, using a definition similar to our own: "context-independent words which occur with high frequency across disciplines" (Inman 1978, p.246). Without having all of her data to hand, it is difficult to account for her much higher figure for semi-technical vocabulary (notice that her figure for technical vocabulary is very close to our own), except to speculate that perhaps a more rigorous selection procedure was applied in the case of the present study. Some words which are used in a variety of scientific contexts but which occur with a specialised meaning in electronics were listed under the heading of technical vocabulary, for example *base, resist, field, potential, frequency, positive, negative*, etc. (see 3.6 below for a further discussion of this point). In addition, Inman's third category was "function words" rather than general words. Many of the words which on intuitive criteria would seem likely to be encountered in basic general English courses would probably have been included under Inman's "sub-technical" vocabulary: *would, should, close, open, important, difficult*, etc.

In order to validate the results of these investigations of semi-technical vocabulary in the two senses, i.e. globally, and within a particular subject area, and to discover what the semi-technical vocabulary of electronics English might look like in relation to semi-technical vocabulary generally, the LOB semi-technical list and the electronics English semi-technical list were compared. It should be noted that the LOB list consists of all semi-technical graphic words which occurred more than 20 times in 160,000 words, or with a greater than 0.0125% frequency of occurrence, while the electronics list consists of all semi-technical lemmas which occurred more than 5 times in 20,000 words, or with a greater than 0.025% occurrence. The fact that the list which is lemmatized, i.e. the electronics list, is also the

list which has a higher percentage frequency of occurrence as a lower cut-off point should make the lists roughly comparable, but it is acknowledged that this might not be entirely satisfactory from a statistical point of view.

Out of the total of 467 graphic words in the LOB Corpus semi-technical list, 215 occurred in the electronics English alphabetical list, and are therefore part of the semi-technical vocabulary of electronics English, by definition. In addition, 17 graphic words on the LOB semi-technical list had closely related graphic words on the electronics English alphabetical list (e.g. *allowed* on the LOB list and *allow* on the electronics list), making a total of 232. These 232 graphic words are all asterisked in the LOB Corpus semi-technical list (Appendix A), so that graphic words on this list not found with any significant occurrence in the electronics English corpus may be seen at a glance. In some cases there does not seem to be any reason for supposing that such items could not occur in electronics English text, for example such items as *particular*, *certain*, *methods*, *main*, *problem*, *total*, *several*. In the case of a number of other items, however, we can probably say by introspection that they would indeed be unlikely to occur in electronics material: *influence*, *individual*, *behaviour*, *knowledge*, *evidence*, *associated*, *sense*, *idea*, *modern*.

If we now examine the electronics English list, which consists of 224 lemmas, we find that 133 lemmas have related graphic words on the LOB list. Again, each of these lemmas is asterisked, so all lemmas which are part of the semi-technical vocabulary of electronics English but are not among the most frequent semi-technical items in science generally may be seen (Appendix D). Here we have a number of words which seem generalizable: *operate*, *demonstrate*, *alter*, *usual*, *illustrate*, etc. and a number of words which seem characteristic of electronics: *connect*, *emit*, *transmit*, *peak*, *repel*, *store*, *contact*, etc.

So, while there are significant correspondences between the two lists, there are also clear differences, the most important being the fact that 91 semi-technical lemmas from the electronics English list, many of them high in frequency, are not prominent semi-technical items in science generally. Carter notes that "different subjects ... have their own lexical cores" (1987, p.185). This fact highlights the relevance of lexical research in individual subject areas to cater for the needs of students of individual subjects.

3.5 Description of the data

The corpus yielded a total of 1,258 lemmas out of 20,017 words of

running text. Lemmas were assembled in rank order, position in the list being determined firstly by frequency and secondly by range. Thus two items with 76 occurrences which were both found in 9 texts would have the same rank, but one with a frequency of 76 and a range of 8 would rank one position lower. All lemmas representing a total frequency of more than 5 in the corpus are listed in Appendix C, and are given detailed treatment in Farrell 1989 (Appendix E), with individual listings for each graphic word included under a lemma. Lemmas representing an occurrence of less than five tokens are not listed.

The first seven words on the list look familiar: *the, is, a, of, to, and, in*. These are the short grammatical words which are usually seen at the head of frequency lists, and which show an extremely high frequency of occurrence. As Sinclair & Renouf put it, "The common words are very common indeed" (1988, p.154). In this study these seven words make up 5,674 occurrences, or fully 28% of the text. The word *the* alone accounts for over 10% of the corpus. A comparison with the first words in the Birmingham Corpus frequency list (which is a survey of general English) shows that with the exception of *is* (which is broken up into its various forms in the Birmingham Corpus) these first items correspond, though the order is different: *the, of, and, to, a, in*. However, the lists then begin to look rather different - some technical items appear very high on the frequency list in electronics English: *current, volt, circuit, resist, charge, electron, coil*, intermingled with further function words: *it, this, by, as, with*. By contrast, the Birmingham Corpus list is dominated by low information grammatical words; the first word referring to a concrete item is people, occurring at position 72 in the list (Sinclair & Renouf 1988, p.149). Sinclair & Renouf note that "Only *time, people, new, know, man* and *little* bring any great semantic content into the top hundred" (1988, pp.154f.). However, the grammatical words seem to have a higher frequency of occurrence in technical language; the first ten words on the Birmingham Corpus list together account for 17% of the corpus (Sinclair & Renouf 1988, p.154), by contrast with 28% for the first seven words in electronics English.

3.6 Syntax

A pedagogical word list which failed to record the syntactic features of occurrences of words would be impoverished. A listing for the graphic word *produced*, for example, could represent an occurrence of the passive, the present perfect, the past simple, an adjective or a hypothetical condition (see 3.9 below). Though our main emphasis in

this study is on lexis, from a pedagogical point of view we would certainly be interested in the relative frequencies of occurrence in different syntactic patterns of forms like *produced*. In the detailed word list, therefore, (Farrell 1989, pp.132-202) grammatical information has been recorded for almost every graphic word. This data, then, is also a useful record of the frequency of occurrence of syntactic patterns in the corpus.

A number of grammatical features distinguish special from general language, among them the high frequency of the passive. Sager et al. note that "extensive use of the passive is one of the most prominent characteristics of technical writing" (1980, p.209). In our detailed list the frequency of occurrence of the passive may be easily checked by a glance at the entry for the lemma *be*: the various forms of this word as auxiliaries, mostly *is* and *are*, are found in passive constructions 456 times (Farrell 1989, p.132). In addition, 114 further occurrences of the passive are found in reduced relative clauses of the type: *the amount of energy stored in the magnetic field*, where both the relative marker and auxiliary are deleted. According to Sager et al., "The desire for conciseness of expression means that relative clauses in SE are very often reduced" (1980, p.223). Modal passives are also relatively common: of the 101 occurrences of the graphic word *be* as a passive auxiliary, 89 are in constructions of the type: *currents will be induced*, and the remainder are accounted for by passive infinitives, e.g. *to be replaced by a piece of wire*. Occurrences of the present continuous passive, e.g. *electricity is being produced*, may also be identified from the entry for *is* and *are* (Farrell 1989, p.132); only 2 are found in the corpus. But the present continuous passive is a popular structure in ESP coursebooks; *First Course in Technical English, Book 2*, for example, devotes two pages of exercises to this syntactic pattern (Beardwood et al. 1979, pp.120ff.). This again points up the relevance of research and the weaknesses of native speaker intuitions in designing syllabuses and materials.

Most occurrences of the past participle are accounted for by the various forms of the passive mentioned above, but also significant are occurrences of the past participle as an adjective (87 occurrences), e.g. *charged sphere*, the most prominent words used in this way being *charged* (15 occurrences), *closed* (8), *induced* (8), *integrated* (6). *Given* as an adjective is found twice, e.g. *a given potential*, with the meaning of "certain". In addition the past participle is found in present perfect constructions 21 times. The *-ed* form of the verb is also found in the past simple (6 occurrences) and hypothetical conditions (4

occurrences).

The present participle has a relatively high frequency of occurrence in the corpus: it is found 317 times in all. Of these, the majority of occurrences are verbal, though not the present continuous tense. The present participle often follows *by* to express the means of doing something, e.g.: *field strength is increased by increasing the number of turns*. This may also be expressed without the use of *by*, e.g. *compare the strengths of the magnetic fields, using the same 2m. length of copper*. Widdowson notes that in scientific writing the present participle may combine the expression of simultaneous action with the expression of resultant action; he calls this causal co-occurrence (Widdowson 1979, p.59). An example from the electronics English corpus would be: *The figure shows how current flows from the capacitor, producing a field in the coil*. The most common present participle is *using* (31 occurrences), followed by *making* (9), *measuring* (8), *producing* (6), *connecting* (6) and *changing* (6). The present participle is found as an adjective or premodifier more frequently than the past participle, with 112 occurrences, for example *reversing switch, varying voltage, connecting wires*. The most frequent present participle premodifier is *alternating* (39 occurrences; probably universally collocating with *current*), then *varying* (10), *oscillating* (9), *following* (9; this is a cataphoric device), *moving* (7). *The following* also occurs as a gerund to refer forward in the text, twice. Some *-ing* forms are conventionally nouns, e.g. *reading* (20 occurrences), *heating* (2), *building* (1). *Providing* makes one appearance, as a conjunction meaning *if*. The present participle is also found in reduced relative clauses (20 occurrences), for example: *currents oscillating at other frequencies*. The present continuous tense occurs only 12 times.

As we noted above, a single graphic word may represent a number of different syntactic patterns. Equally, to say that a certain syntactic pattern is frequent in the corpus does not mean that it will be evenly distributed lexically. For example, in the case of the lemma produced (overall frequency 79) there are 17 occurrences of the third person singular present simple *produces*, while in the case of *connect* (overall frequency 68), this form does not occur at all. Conversely, in the case of *connected* there are 20 occurrences of the imperative form, while this form is not found at all with *produced*.

A notable feature of the occurrence of nouns in the corpus which is particularly relevant from a lexical perspective is the fact that a high proportion occur in noun/noun constructions of the type *energy transfer*, where the first noun has an adjectival function. (In some

cases, a noun and adjective share the same form, e.g. *potential*.) Occurrences of such nominal compounds may be identified in the detailed list (Farrell 1989, pp.132-202) where nouns are marked adj.; there are exactly 600 such occurrences in total, making this feature more frequent than the passive, and therefore the most salient syntactic characteristic of the electronics material. Sometimes the frequency of a noun as an adjective in nominal compounds is higher than its frequency as a noun, e.g. *input* occurs 29 times as an adjective or "premodifier" (Leech & Svartvik 1975, p.271), but only 9 times as a noun. Sager et al. distinguish between three types of what they call "determinants" in nominal compounds: "... those which designate objects, those which designate properties, and those which designate processes and operations" (1980, p.268). Applying this analysis to the electronics English corpus, a number of these different types of noun compounds may indeed be identified, including a few not mentioned by Sager et al.

Even within the first group, i.e. noun compounds in which the determinant designates an object, the function of the determinant seems to vary, as noted by Sager et al. (1980, pp.268-71). For example, in one type the determinant denotes the material of which an object is composed (Sager et al. 1980, p.269). Examples from the electronics English corpus would be: *copper wire*, *zinc plate*. In a related type the determinant describes what an object is composed of, though not a concrete material: *electron beam*, *sound wave*, *radio wave*. The determinant may also designate the precise function of a piece of equipment: *current amplifier*, *voltage amplifier*, *radio transmitter*. This last type seems to be particularly common in the electronics material, on a purely informal inspection. And finally, the determinant also seems to have the function of denoting a piece of equipment of which the nucleus (or second noun) is a component part (Sager et al. 1980, p.270). Electronics examples are: *aerial wire*, *compass needle*, *diode valve*, *emitter resistor*.

In Sager et al.'s second group the determinant specifies the object of which the nucleus is a property (1980, p.271). Examples of this type of compound from the electronics material might be *collector current*, *base voltage*, *gate potential* (*collector*, *base* and *gate* are electronic components). Their third group comprises compounds in which a process is described (Sager et al. 1980, p.271). Electronics examples are *charge flow*, *current gain*, *energy transfer*. Sager et al. note that confusion may arise in interpreting such compounds in relation to transitive/intransitive features of the verb underlying the nucleus,

giving the examples *temperature change* and *temperature control*. In the first case it is the temperature which is itself changing (subject), while in the second the temperature is being controlled (object). Similarly, it is the charge which is flowing in *charge flow*, while the energy is being transferred in *energy transfer*. Confusion for learners could also arise between compounds of the type *current amplifier* and *collector current*, the learner perhaps assuming that *collector current* means a current which collects something, by analogy with the first. An English teacher could perhaps find a role here, though most noun compounds in special English are decidedly technical.

3.6 Polysemy

The idea of a general language of science, or semi-technical vocabulary, has been discussed above. Most writers who have investigated this notion seem to be talking about words which are part of the general lexicon and which retain their general meaning when used in special texts: *obtain, occur, require, produce*, etc. Trimble has defined "sub-technical vocabulary" as "context-independent words which occur with high frequency across disciplines" (1985, p.129). But Trimble then goes on to extend the definition of sub-technical vocabulary to include "those common words that occur with special meanings ... words that have one or more general English meanings and which in technical contexts take on extended meanings" (*ibid.*). A number of such items were noted in the present study, but it was felt that they belonged more properly under the heading of "technical vocabulary", as they usually seemed to require a precise technical explanation by the subject teacher.

However, if the general and technical senses of an item are clearly related, or polysemous (Jackson 1988, p.5), this may be of assistance to ESP learners, and the English teacher could play a role in bringing out the connection between general and technical meanings.

If we look at the list of technical words from the study of electronics English in Appendix D, a number of instances of polysemy may be seen. For example, *current* in the technical sense of a flow of electricity may be linked with its general reference to a flow of water. The general meaning of the word *resist* may be linked to its technical sense of obstructing, or slowing a flow of electricity. *Capacitor* may be explained by its "capacity" to store electricity. Other similar examples might be: *circuit, coil, potential, energy, frequency, wave, induce, terminal, insulate, transformer, generate, earth, junction, gate, series*. In fact, very few of the words on the technical list are really "technical"

in the sense of being completely unknown in general language: *anode, cathode, diode, thermionic, solenoid, ammeter, silicon, voltmeter, reactance, ohm, semiconductor, electromagnet, amp, galvanometer, electroscope, rheostat, oxide, flywheel, coulomb, impedance, dielectric, sinusoidal, triode*; this is the complete list.

However, English teachers might find themselves on tricky ground here, and some may prefer to avoid dealing with these items altogether. In some cases, a related general meaning may give only a vague idea of the technical meaning, which may then need to be stated in very precise terms: the word *energy*, for example. Sager et al. refer to this as "redefinition": "General notions are often the basis for scientific investigations, but ... special reference then requires the definition of the concept .... Such a definition usually entails reducing the extension of the general language word used" (1980, p.255). They give the technical definition of the word *force* as an example (this word occurs in our list of technical items): "*force*: a force is any influence that can cause a body to be accelerated" (1980, p.255). They contrast this definition with that of the *Concise Oxford Dictionary*, which defines force as "strength", "power", "impetus", "violence", "intense effort", among others. Clearly, it is the precise technical definition which is the one of interest to learners; the subject teacher would not give many marks to a learner who defined *force* as "intense effort" in an electronics examination, and even fewer if he defined it as "power", which has a separate precise technical meaning of its own.

3.7 Homonymy

In some cases, an English teacher might assume a polysemous relationship incorrectly, for example the word *base* may not be used in its general sense of "bottom", or "supporting part" in technical language. It is defined in electronics as "Part of a valve where the pins that fit into holes in another electronic part are located", or "The middle region of a transistor" (Trimble 1985, p.130). This, of course, is homonymy, which "refers to words with different meanings sharing the same form" (Jackson 1988, p.5). Dealing with this type of item may be beyond the competence of the language teacher, but he should be aware of items used in general language which have an unrelated technical sense, without necessarily being *au fait* with the precise technical definition. Generally, however, most technical items seem to be either purely technical (see list above), or to show some polysemy with general language, however tenuous.

Homonymy is seen in some of the general and semi-technical

items. The word *make*, for example, occurs 51 times with the sense of "to create" or "build", but also 11 times with the sense of "to cause something to happen". *Close* is found 5 times as an adjective meaning "near", and 4 times as an imperative verb meaning "shut". *Set* occurs 7 times as a verb with the sense of "to put controls at a certain position", 8 times as a "radio set", and 4 times as a noun meaning "a group of items". *Above* is seen as a preposition denoting spatial relations 9 times, as an anaphoric word referring to earlier parts of the text 9 times, and with the meaning of "more than once". *Form* is found as a noun meaning "shape" 7 times, as a verb with the sense of "create" 7 times, and as a noun meaning "type" 4 times. The word *case* occurs with the sense of "eventuality" on 14 occasions, but with the meaning of "outer covering" or "casing" 4 times. *Common* is seen with the sense of "usual" 8 times, and as an adjective meaning "in both" (e.g. *common base*) 7 times. *Take* occurs in its ordinary sense 10 times, as a verb denoting how much time an event occupies 6 times (e.g. *the spot takes 10m/sec. to ...*), and as a passive verb meaning "is presumed" on 2 occasions (e.g. *X is taken to be Y*). *Unit* is found 3 times referring to a piece of equipment, twice in the collocation *a unit of measurement*, and twice referring to a chapter or sub-division of the textbook.

3.8 Anaphora

A glance at the beginning of the detailed list (Farrell 1989, pp.132-202) shows that anaphora is prominent in electronics English: one of the most frequently occurring items is *it*, with 153 occurrences, 134 of which are anaphoric. (The remaining 19 occurrences of *it* are also of interest; see 3.9 below). Occurrences of the word *this* are divided almost equally between anaphoric reference (84 occurrences) and use as a demonstrative adjective (89). Other words used anaphorically are less frequent. Perhaps less predictably, the word *one* is next in order of frequency, with 13 anaphoric occurrences, e.g. *two coils of wire, like the ones you have been using*. The word *above* is used to refer back in the text or to a previous diagram on 8 occasions, 5 times as an attributive adjective, e.g. *in the above circuit*, and 3 times as a prepositional adverb (Leech & Svartvik 1975, p.87): *the lines shown in the patterns above*. The word *that* is used anaphorically 6 times, for example: *a charge equal to that on an electron*. The word *another* also occurs 5 times anaphorically, e.g.: *if the transistor is changed, even if it is replaced by another of the same type*. And finally, the word *each* occurs with anaphoric reference 4 times, for example: *Make a list of electrical appliances in your house. What is the most important*

electrical effect that each makes use of?

3.9 Collocation

As noted above, some limited collocational studies were carried out on the electronics material. The first was a slightly more in-depth examination of recurrent patterns of use of a few low-information grammatical words than is seen in the detailed list for the electronics material. In discussing the COBUILD project Carter notes that "Information about the operation of ... content-less grammatical words ... in their normal discourse environments ... is invaluable" (1987, p.142). The words examined were *in*, *as*, *that*, *at*, *one* and *it*. The data produced from this special examination is found in Farrell 1989 (Appendix F, pp.203-206), and all these items are asterisked, with a reference to the appendix, in the detailed list (Farrell 1989, pp.132-202).

Most occurrences of *in* are predictable: as a preposition, denoting spatial relations (132 occurrences). A large number of occurrences fall into the pattern *increase/change in voltage/current/direction*, etc. (68 occurrences). Fixed expressions account for a large number of occurrences: *in this case* (13), *in series with* (7), *in a circle* (5), *in this way* (3), *in effect* (3), etc. (see Farrell 1989, Appendix I for further examples). Also of significance is the use of *in* with the meaning of "while doing" something: *errors occur in the measurement of* (7), *in experiments* (7), *in the insulation of cables* (1), *in changing a.c. into d.c.* (1).

As is most frequently used with the meaning of "like", "similar to", or "in the manner of", e.g.: *used as a voltage amplifier*, *as shown*, *as above*, etc. (74 occurrences). Also common is its use meaning "while", relating change in one variable to change in another, e.g.: *lines move outwards as the current changes* (27 occurrences). "As .... as" constructions account for a number of occurrences, e.g. *as far as possible*, *as closely as possible* (16 occurrences). Also important is the use of *as* with the meaning of "because" (12 occurrences).

That is most frequently found as a conjunction following verbs, particularly *show*: *shows that*, *note that*, *assume that*, *check that*, etc. (63 occurrences). A number of occurrences express purpose, following *so* (23). The anaphoric uses of *that* are of interest (see 3.8 above).

Also used anaphorically is *one* (13 occurrences; see above), though its most frequent use is as a numeral (19). 14 occurrences are in *one ... the other* constructions. Most occurrences of *at* relate to a measure of voltage, temperature, etc., e.g.: *is held at 6v* (45), while its frequency as a spatial preposition is lower (29). Finally, a number of

uses of *it* which are not anaphoric were noted, e.g. *it takes the spot 10m/sec. to ..., it should be remembered that ....*

In addition, three lemmas were selected for closer scrutiny: *produced, make, connected*. In this case fuller collocations are given; every line from the electronics English corpus in which these items occurred is presented in Farrell 1989 (Appendix G, pp.207-212) with the preceding and following context. (In some cases two lines are included where one would have been insufficient to illustrate the use of the item in a particular context). A glance at this data shows that the immediate collocates of semi-technical items are often technical. *Produced* and *connected* were selected because they are two of the most frequently-occurring semi-technical items. *Make*, though defined as a general word, was selected because it seemed to show patterns of use perhaps not often dealt with on general courses, and for the purpose of contrast with *produced*, of which it is sometimes a synonym.

Of the different graphic words subsumed under the lemma *produced*, the word *produced* is the most common, by definition (30 occurrences). Of these occurrences, *produced* is most often found in reduced relative clauses (see 3.6 above), often preceded by *magnetic field* and followed by *by*, for example: *the magnetic field produced by the same piece of wire* (19 occurrences). Full passives account for 9 further occurrences e.g. *electricity is produced in varying amounts*. The graphic word *produce* is found 21 times, 9 times following modals: *it can produce electricity*, 4 times as a purpose infinitive, e.g. *their purpose is to produce a uniform magnetic field*, 4 times as a plural verb, and 4 times as an infinitive following an adjective, e.g. *it is still possible to produce a distorted output*. The 17 occurrences of *produces* are as singular verbs. More interesting is *producing*, which shows 5 examples of Widdowson's "causal co-occurrence" (1979, p.59; see 3.6 above), e.g. *collapses very rapidly, producing such a large rate*.

The most frequently occurring meaning expressed by the various forms of *make* is "cause" (48 occurrences) rather than that which would intuitively spring to mind: "create" (20 occurrences). The remaining occurrences are accounted for by fixed expressions and collocations: *make use of* (2), *make up for* (1), *make up* (i.e. "create") (1), *make a list* (1), *make adjustments* (1), *calculation can be made* (1), *connection may be made* (1), *making connections* (1), *making measurements* (1). Among the uses of the graphic words *make, made, makes, making* with the meaning of "cause", the change caused may be expressed by a verb, adjective or noun in the following clause. The

most common is a verbal construction, found 16 times following *make*, e.g.: *to make it oscillate continuously*, 5 times following *makes*: *this makes a current flow*, 3 times following *making*, e.g.: *making the photo cell current vary*, and twice following *made*, e.g.: *the coils are made to rotate*. Adjectival constructions are also common, following *made* 11 times, most commonly with the adjectives *positive* and *negative*, e.g.: *If the base voltage is made positive*, following *make* 5 times, e.g.: *negative charges nearby make it easier to ...*, and following *makes* 3 times, e.g.: *which makes the solid state diode possible*. Finally, a noun is sometimes used, once following *make*: *electrons thus produced make the valve a very good conductor*, and once following *makes*: *time-base switched off makes an appropriate voltmeter*.

In the case of *connected* the third person singular present simple form *connects* does not occur at all. Note the contrast with *produced*, where it occurs 17 times (see 3.6 above). The most common graphic word is *connected* (24 occurrences), 14 of which are full passives, e.g. *when the grid is connected to the filament*, and 10 reduced relative clauses, e.g. *the earphone connected to the rectifier*. *Connect* occurs 22 times, almost always as an imperative verb (20 occurrences). The noun *connection* occurs in singular and plural forms 13 times, 4 times as the nucleus of a noun compound (see 3.6 above), e.g.: *all the battery connections are reversed*. The present participle *connecting* is found 9 times, on 6 occasions to express the means of doing something, often following *by*, e.g.: *current can be smoothed by connecting an inductance coil*, twice as an adjective, e.g.: *the connecting wires*, and once in a reduced relative clause: *wires connecting where they cross*.

3.10 Affixation

Where there is an emphasis on lexis in materials, a great deal of attention is often paid to affixes, and prefixes in particular. Wallace proposes a list of prefixes as deserving of attention, including the following: *un-*, *de-*, *dis-*, *mis-*, *mal-*, *sub-*, *pre-*, *semi-*, *trans-* (1982, p.87), an approach which seems promising in helping learners to predict the meanings of unknown items (Nuttall 1982, pp.67f.). However, a check of the data proves disappointing; surprisingly enough, only three of the prefixes Wallace mentions occur in the alphabetical list for the electronics corpus: *semi-*, *trans-* and *re-*. *Semi-* occurs in only one word, *semiconductor*, with a frequency of 13. *Trans-* occurs in *transfer*, *transform*, *transistor*, *transmit*, though it seems unlikely that its semantic function in these different words can be linked in any way

which is useful pedagogically. *Re-* occurs commonly, e.g. *remove, repeat, repel, replace*, etc., but in this case too the semantic function of the prefix is buried too deep in the etymology of these words to be of any use in linking these items for learners. The only productive prefix in the corpus seems to be *therm-*, occurring in *therm, thermal, thermally, thermionic, thermionics, thermocouple, thermoelectric, thermometer, thermostat*, all items connected with heat. Also of note is *over-* in *overheat, overcome, overlap*. More important seem to be suffixes, which generally have a grammatical rather than a semantic function, e.g. *-s, -ed, -ing, -ion, -ic*. Awareness of these may be developed through the use of word building tables (see section 4).

4 Lexis in ESP materials design

We now come to the practical applications of this study, and an examination of how the semi-technical word lists and other lexical data produced by the study of the English of electronics and the LOB Corpus might be used in designing materials for ESP courses, whether common-core general-purpose ESP courses or subject-specific courses aiming to cater for the needs of students of a particular subject specialism.

4.1 The common-core approach

We noted in section 1 the importance of the common-core approach in defining the role of the ESP teacher. Johns and Dudley-Evans (1980, p.140) argue in favour of a common-core approach: "Since 1971 the English for Overseas Unit at Birmingham University has concentrated ... on the "common-core" language problems of students (including vocabulary). The experience of the past seven years has confirmed that the "common-core" approach has both theoretical validity ... and practical applicability". They note, however, an inevitable difficulty in designing common-core courses: "a common-core programme ... entails a degree of abstraction". In other words, how do we focus on common core syntactic and lexical features of science without using texts from specific areas of science, and without this approach degenerating into explanations of grammar rules and learning lists of words?

The difficulties involved in designing common core courses may be seen in some existing coursebooks, *Nucleus* (Bates & Dudley-Evans 1976), for example. The first book in the *Nucleus* series is called *General Science*, and is followed by a number of specific coursebooks

for different branches of science and technology. In writing about the design process involved in *General Science*, Bates (1978, p.2) underlines the importance of semi-technical vocabulary: "*General Science* presented the common-core language of scientific and technological description, including items of general use which were important to the scientist - e.g. *consist, depend, relationship* .... This ... might well constitute greater obstacles to understanding than the more obvious technical vocabulary." However, an examination of the material in *General Science* and the word list at the back of the book shows that while semi-technical items are prominent, there is quite a high proportion of use of subject-specific technical items from different subject areas: *anhydrous, cartilage, aorta, iridium, cadmium, epithelial, cytoplasm, auricle, glycogen*, for example. The technical vocabulary of a particular subject is relevant to students of that subject, but it seems unhelpful to ask learners on a common core course to cope with technical vocabulary which is from a subject specialism they will never study. This is also a failing of other attempts at common-core ESP courses. Very often "common-core" seems to mean in effect a succession of units devoted to different subject areas, with an emphasis on grammatical features of science generally. For example, *First Course in Technical English, Book One* (Beardwood, Templeton & Webber 1978), which is a coursebook for technical students generally, begins well with general units on tools, materials, containers, instruments, shapes, motor vehicles and instructions. Then there is a unit on basic electricity which is probably more or less accessible to the layman, whether learner or teacher. But the next two units of the book consist of rather "heavy" technical material on the properties of engineering materials, with prominence given to such subject-specific technical items as ductility and malleability. This material is formidable and off-putting to students of a technical area other than metallurgy. Understanding the material in a common core course should not involve grappling with subject-specific vocabulary from an irrelevant subject specialism; the learners would surely be better off becoming familiar with subject-specific vocabulary which is at least relevant to their studies.

The basic problem with such courses is the fact that they concentrate on the common core syntactic features of science at the expense of common core lexical features. We have identified a list of items which may be said to represent the general vocabulary of science, i.e. the LOB Corpus semi-technical word list (Appendix A). In what specific ways, then, might such a list play a role in developing a common core

course for ESP learners? The following are some suggestions:

(a) As a reference list for materials writers to consult when designing specially written texts. In spite of the recent widespread interest in using authentic material in ELT generally, there still seems to be a role for specially written material, particularly in ESP, where authentic material usually means subject-specific material from the subject textbook (itself specially written for science students) or other sources. Hutchinson & Waters feel that in an ESP setting "there is ... no intrinsic merit in an authentic text", and suggest that what is more important is what role one expects a text to play in the learning process. They go on to point out that if one wanted to carry out, say, a jigsaw reading task, it could be virtually impossible to find a set of appropriate authentic texts (1987, p.159).

The materials writer could try to give greater prominence in materials to words high on the list in terms of frequency and range, without perhaps needing to be overly precise about this. Because technical vocabulary has been excluded, the materials produced will not be "scientific" in the strict sense, but could be similar to general interest texts of the type found in newspapers and magazines: popular psychology, the paranormal, nature, the environment, climatic trends, trends in social habits; all of these text types will be found to contain a high proportion of semi-technical vocabulary.

An advantage of using this type of material on ESP courses is that it can provide a refreshing antidote to purely technical material. Hutchinson & Waters argue that "in terms of language content there is little justification for a subject-specific approach to ESP" (1987, p.166), and that we should "make learners aware of the lack of specificity of their needs" (1987, p.167). ESP teachers are sometimes overly concerned that their materials should be relevant, i.e. subject-specific, in the belief that learners will be more motivated by this type of material. This may not be so; some learners may have motivational problems in relation to the subject itself. Even those learners who are highly motivated to learn the subject may welcome a break from technical material, which always requires concentrated effort. Also, it is extremely difficult for the English teacher to make technical material communicative; only the subject teacher working with students in the technical laboratory is in a position to bring the language of the subject specialism to life, a point made by Hutchinson & Waters: "The [technical] teacher does not just give a lecture about a machine, he uses the actual machine or a model of it as an integral part of the delivery" (1981, p.60). Semi-technical materials of the type suggested above, however, may serve as useful

stimuli for animated class discussions.

(b) As a checklist with which to evaluate the possible usefulness of authentic materials, again of the semi-technical type suggested above. The chief difficulty here might be the fact that all texts contain a high proportion of rare words, and it could be argued that from a lexical point of view this is then no improvement on using scientific texts from an unfamiliar subject specialism. However, unusual items acquired from semi-technical texts might turn out to be useful to learners at some point, unlike the highly technical vocabulary of another subject.

(c) Learners could themselves be given the list, together with the information on frequency and range, with the caution that the teacher is not putting them under pressure to absorb all the items on the list, but that it may be used as a reference tool. If they are worried about the meaning of an unfamiliar item in a text, for example, they could be encouraged to use an alphabetical version of the list to check whether the word is an important one in science generally. This would be in line with recent interest in making learners privy to information usually available only to teachers, and encouraging them to take responsibility for their own learning.

(d) In the design of a lexically based course. We began to explore the idea of a lexical syllabus in section 3 above. However, we have not yet examined what materials having lexis as an organising principle might look like. Carter & McCarthy note that "A coursebook based on this [i.e. lexical] syllabus would be radically different from conventional ones" (1988, p.160). Very few attempts have been made to date at designing this type of course. A notable feature of those few attempts is that where a list of words is the organising principle, the course tends to look like an elaborate dictionary, or encyclopaedia.

4.2 The subject-specific approach

Often an ESP teacher, rightly or wrongly, will come under pressure from learners and/or administrators to implement a subject-specific approach. Hutchinson & Waters note that "learners often demand subject-specific texts" (1987, p.166). We have discussed the difficulties that the English teacher may have in exploiting technical material. If there is a demand that the English course be directly relevant to the learners' technical studies, then this requirement may be answered by focusing on the lexical characteristics of their subject of study. We have seen that not only may special English be primarily distinguished from general English by lexical rather than syntactic features (Johansson 1975, p.1), but also that individual subject areas are characterised by

distinct non-technical lexical features. In section 3 we found a list of items with varying degrees of prominence in electronics English which were semi-technical, but which did not occur on the LOB Corpus list of the general vocabulary of science. Lists of semi-technical items characteristic of individual subject areas may then be used by the English teacher or course writer in ways similar to those suggested above: in writing special texts, in evaluating authentic material, as a reference tool for learners, and in developing subject-specific lexical courses to supplement common core materials, without having to use specialised materials or explain subject-specific technical vocabulary.

As well as identifying semi-technical items which can play a role in developing materials accessible to both teacher and learner, other aspects of the data gathered from lexical studies of special registers are invaluable in subject-specific materials design: for example, the relative frequency of occurrence of different semantic values of homonyms may be identified (see 3.7 above), as well as recurrent collocational patterns which may provide evidence running counter to intuitions. For example, in the electronics corpus the word *into* occurs 4 times in the pattern *changes electricity into heat*, and not at all in its "typical" or "basic" context of movement *into* a container, closed area, etc.

4.3 Existing lexically-organised courses

An early and very interesting attempt in this direction is not in fact a course, but a reference book: *A Reference Book of English Words and Phrases for Foreign Science Students* by R.F. Price (1975, first published 1966). However, as we have suggested above, perhaps a lexically-based course is inevitably a type of specialised dictionary. Price himself seems to be hinting at a role wider than merely reference in his foreword to the book: "This ... book ... is not intended to *teach* science, or scientific terms, but the non-technical English words and phrases which are necessary to describe and explain things and events scientifically" (1975, p.vii; my italics).

This body of material, then, has semi-technical vocabulary as its organising principle, and so is of particular interest to this study. Price gives the following instructions to users of the book: "When using this book you may start from two places: (i) If you have a word you want to understand, begin at the index at the back. (ii) If you have an idea which you want to express in English, begin at the list of contents at the front" (1975, p.ix). The list of contents at the front is arranged in notional categories: *shape*, *composition*, *texture*, *colour*, *taste*, *smell*, *quantita-*

tive relations and size, spatial relations and order, temporal relations, whole and part, change, motion, techniques and use of apparatus, facts, concepts and problems, scientific method, the particular, the general and comparisons, causation, classification.

In each section of the book, headed by one of the notional categories above, a list of words is first given, e.g. under the first category of *shape* are listed the words *shape, form, line, angle, straight, curve, parallel, circle*, etc. Inflectional and derivational forms are given in the case of each word, in the following order: verb infinitive, past tense, past participle, noun, adjective, adverb. So, a typical entry looks like this: *to circle, circled, circled, circle, circular.* Where one of these forms is used with a different meaning, or does not exist, as with the adverb form of *circle*, a blank appears. There are a few obvious flaws in this presentation system. Price could not have carried out a serious analysis of information which has long been available about the relative frequency of occurrence of different inflectional forms in scientific writing. In the electronics English corpus there were only 6 occurrences of the past simple, but 317 occurrences of the present participle, yet Price has included the former and omitted the latter. Neither is any attention paid, for example, to the frequent use in science of nouns as premodifiers in nominal compounds (see section 3 above).

In each notional category the word list is followed by a section exemplifying each word in sentences, usually accompanied by a diagram to make the meaning clear. In some cases, all the forms of an item in the word list are exemplified, e.g. *This is a sphere. This object is spherical.* In other cases, not all forms are exemplified, e.g. *to spiral, spiralled* (past tense), *spiralled* (past participle), *spiral* (noun), *spiral* (adjective), *spirally* are found in the word list, but only an example of *spiral* as a noun is given: *This is a spiral* (picture supplied).

Synonyms are exemplified by demonstrating their interchangeability in sentences: *This container/box has a cover/lid which shuts/closes when placed over the top.* However, no attempt is made to exemplify contexts where synonyms might not be interchangeable, e.g. in electronics we would talk about a *closed circuit*, but not a *shut circuit.* Homonyms are simply disregarded; in our own study of electronics English *close* was found to occur 5 times as an adjective meaning "near", and 4 times as an imperative verb meaning "shut".

Skills for Reading, a collection of reading materials developed at Reading University by Morrow (1980), seems to address some of the issues we have been raising in relation to the role of the ESP teacher.

The reading texts presented in the book are abridged but entirely authentic materials taken from *New Scientist*. In his introduction to the book, Morrow notes that articles in *New Scientist* are designed to be "accessible to the non-specialist... educated layman" (1980, p.9). This means that the material can be used by non-specialist English teachers in the classroom. Morrow continues: "The role of these texts is to provide a vehicle of general interest for the practising of reading skills which can subsequently be applied to whatever texts the student wishes to study" (1980, p.10).

A review of the materials in *Skills for Reading* shows a high proportion of semi-technical items. A great number of technical and context-specific items also occur, but the point is that these materials are useable by an English teacher; they have been written by professional journalists for the layman and are therefore stimulating and provide good talking points for related communicative activities. There are, for example, articles on the depletion of the ozone layer, the biological clock, fireballs, homes for refugees, landing on Venus. In addition, many of the related exercises in *Skills for Reading* are lexically-oriented; there are word-building exercises using derivations, e.g.:

exclude	exclusive	exclusion
exceed	.....	.....
.....	adjustable	.....

(Morrow 1980 p.52)

Learners fill in the appropriate derivatives (if they exist) in the blanks. The given forms occur in the reading text. However, the organising principle of this book is not primarily lexical.

A course which is organised on lexical principles, and one which has aroused a great deal of interest, is *The Words You Need* (Rudzka et al. 1981). In their preface Rudzka et al. acknowledge that the inspiration for the book came from the work of Adrienne Lehrer on semantic fields and lexical structure (1981, p.i). The table of contents is made up of groups of words, organised in semantic fields (e.g. *conceit, pride, self-esteem, vanity*) and synonymous pairs (e.g. *unpleasant/ghastly, unwilling/reluctant, produce/generate*). The groups of words are arranged in grids of two types. One type uses the techniques of componential analysis (Jackson 1988, pp.80-95) to identify the overlapping meanings of synonyms, for example:

	strong	*potent*	*powerful*
producing great force			+
having great force	+		+
having a solid structure	+		
great effect on senses	+		
firm	+		
effective	+	+	
usu. of drink and drugs		+	

(Rudzka et al. 1981, p.28)

The second type of grid brings out the different collocations of semantically related items, for example:

spurn	*decline*	*turn down*	*reject*	*refuse*	
+					*a lover*
+					*sb's attentions*
+	+	+	+	+	*sb's offer*
	+	+		+	*a request*
	+	+		+	*an invitation*
		+	+		*sb's proposal*
		+	+		*an application*
		+	+		*candidate*
			+	+	*a gift*
				+	*sb admittance*
				+	*sb permission*
				+	*sb's money*

(Rudzka et al. 1981, p.47)

In discussing *The Words You Need* Carter (1987, p.170) notes that "Psycholinguistic support for the book's method is ... provided by research reported in Cornu (1979) which indicates that individuals tend to recall words according to the categories or semantic fields in which they are conceptually mapped." Swales supports this view when he notes that learners' "capacity for mental organization should be utilized in lexical work both as a motivating factor and also because there is evidence that schematized data is easier to learn and retain than data that appears arbitrary and disordered" (1983, p.23). Harvey is also enthusiastic about this type of approach: "... students like

playing with words ... particularly when organized on a thematic basis ... it seems likely that people generally find the use of grids and matrices psychologically satisfying" (1983, p.245).

Channell, one of the co-authors of *The Words You Need*, notes that there is evidence from psycholinguistics to suggest that the mind uses semantic similarity in classifying words, pointing out that "slips of the tongue" made by native speakers are often not random mistakes, but produce words from the same semantic field as the intended word. She gives the following example: "I have my book and my jigsaw ... I mean my crossword" (1981, p.117).

Once again this lexically based course tends to have the appearance of a work of reference, in spite of the fact that the word study sections are interspersed with texts. The book's approach is perhaps overly explanatory and expository, not leaving much for the learner to contribute. Swales makes this point about the book, and suggests "getting ... classes to do some of the work" (1983, p.31), emphasising "the role of the LSP learner as an active participant in solving problems of lexical structuring" (*ibid.*, p.32).

4.4 Designing lexically-based ESP courses

The organising principle of any course is the syllabus on which it is based. But a lexical syllabus itself needs an organising principle. We have examined in section 2 and throughout this study the criteria for deciding on a list of items important pedagogically, to be used in designing a lexical syllabus and lexical materials, and we are suggesting that semi-technical words should make up this list in ESP courses. Sinclair & Renouf suggest that the data gathered in lexical research should be itself a lexical syllabus, giving an extract from their lexical syllabus for the COBUILD course for the word *by* which gives very detailed semantic and collocational information about this item running to a page and a half (1988, pp.156-8). This seems too unwieldy a body of information for the title "syllabus", which generally implies making abstractions from language. On the other hand, a simple list of words as a syllabus also seems to leave something to be desired. But if we take a list of words which we have determined are relevant pedagogically, we may then give shape to this list and form it into a lexical syllabus by ordering those items into groups, using notional categories, synonymy or morphology as organising principles. Approaches of this type may be seen in some existing courses. Price organised his words into notional categories of *shape*, *composition*, etc. *The Words You Need* uses synonymy to group words. *The Right Word* (Fowler

1987) has two sections, the first devoted to nouns, which are grouped under notional/thematic headings: *people, shopping, weather, health,* etc., and the second to verbs, which are grouped in synonymous sets, e.g. *gaze, glance, look at, regard, stare, watch*. Morphology is another possible organising principle. This has not been used at the level of syllabus in any existing course to my knowledge, but is used in exercises in *Skills for Reading* (Morrow 1980), and also in *English for Cambridge First Certificate* (Allsop 1979).

Although Johansson's work on the vocabulary of learned and scientific English was descriptive rather than pedagogical, he has made an interesting contribution to the elaboration of a lexical syllabus based on semi-technical vocabulary. In making comments on "open-class words characteristic of SE", he groups these items into a number of notional categories (Johansson 1975, p.22): "... words which ... denote matters which we associate with scientific exposition (*discussion, argument, result, conclusion,* etc.) or procedure (*analysis, experiment, measurement, observations, test,* etc.) ... adjectives descriptive of a scientific approach (*empirical, experimental, objective,* etc) ... statistical terms (*sample, sampling, probability, distribution, random, significant,* etc.) ... classification (*class, group, type, species, items, factors, units, properties, characteristic, typical,* etc.) ... relational words (*same, similar, different, distinct, equivalent, equal, average, normal, relative, difference, increase, change,* etc.).

Taking the list of semi-technical items derived from the study of the English of electronics (together with a few informal items from the general list such as *get*, included for the purpose of contrast with formal equivalents, and technical words for comparison with semi-technical equivalents, e.g. *generate*), I propose grouping items in synonymous and antonymous sets under notional headings, to form part of a possible semi-technical lexical syllabus for an electronics course (a similar approach could be adopted for a common-core course, using the LOB data):

Cause/Effect: *cause/make, alter/change, distort/affect, effect/result*

Similarity/Difference: *same/similar, constant/steady, keep/maintain/continue, differ/vary, alternate/oscillate*

Connection/Separation: *connect/join/contact, attract, remove, replace, transfer, separate, reverse/deflect/repel*

Increase/Decrease: *increase/rise/gain, decrease/lower/reduce/fall/drop*

Relation: depend, relate, proportion, compare

Production: build/construct/make, produce/generate, emit

Measurement: measure, calculate, observe/note/study, accurate, correct, error

- The above is not a comprehensive syllabus, but some suggestions which could be built on. Other synonymous or semantic sets which could be treated together are:

turn/rotate/wind
determine/find/detect/obtain/get
though/although/however
show/demonstrate/illustrate/indicate/represent
usual/common/normal
need/require
around/about
provide/supply
area/region
control/operate
simple/easy
first/primary
act/perform
device/apparatus
instrument/meter
move/slide
part/component
section/piece
vertical/perpendicular
amount/quantity
way/means
bright/fluorescent

Morphology could form the next level of organisation, presenting the various derivationally and inflectionally related forms of an item, e.g. *produce, produces, produced, producing, production,* as in Price (1975). We might consult our data in deciding which forms to list, and which to give priority. For example, the related noun *product* or

adjective *productive* did not occur in the electronics corpus. *Produced*, with a frequency of 30, would seem to merit a greater degree of attention than *production*, with a frequency of 4. However, perhaps simply presenting these forms would miss an opportunity to engage learners actively, as we could by using grids of the type found in *Skills for Reading* (Morrow 1980) and *English for Cambridge First Certificate* (Allsop 1979). Learners could guess, or consult dictionaries or authentic materials to fill in the blanks:

produce	*produces*	*produced*	*producing*	*production*
use	.....	.....	.....	.....
.....	.....	.....	*showing*	.....
.....	*moves*	.....	.....	.....
.....	.....	*connected*	.....	.....
.....	.....	.....	*changing*	.....
measure	.....	.....	.....	.....
.....	.....	.....	.....	*supply*
vary	.....	.....	.....	.....
emit	.....	.....	.....	.....

The semantic and syntactic data in the detailed list for electronics English (Farrell 1989, pp.132-203) would be invaluable in devising exercises to make learners aware of the patterns in which the different related forms of an item are typically used in the target material, particularly if supplemented by full collocational data of the type presented in Farrell 1989 (Appendix G, pp.207-212). Using the data for *produced* (see section 3 above), for example, we might create an exercise using semi-technical contexts along these lines:

Fill in the blanks with *produce, produces, produced, producing, production* as appropriate:

1. In order to grow, plants need the energy ..... by the sun.
2. The ..... of oil in the Middle East has increased rapidly in recent years.
3. A high proportion of smoke pollution ..... by domestic fires.
4. The rubber tree ..... a liquid which is made into rubber.

Or, using the data more directly, subject-specific sentences could be used in the above exercise if desired.

Semantic exercises on the synonymous sets of words could follow the model of *The Words You Need*, incorporating Swales' suggestion of giving the learner the opportunity to contribute to lexical structuring, rather than simply presenting him/her with all the information. Texts or example sentences could be supplied which would exemplify the different semantic features of related items, and the learners could then fill in the grids. Grids may be used to bring out overlapping meanings of items and differences in collocation, though as noted in section 3, most semi-technical items seem to collocate with technical items in the electronics material. Here is an example grid on the set *alter/change/distort/affect* to bring out overlapping meanings (the grid has here been completed by way of illustration):

	alter	*change*	*distort*	*affect*
replace something		+		
cause a change in sthing	+	+	+	+
become different	+	+		
cause undesirable change			+	

The following is a collocational exercise on *produce/generate/make/emit*, using the detailed collocational data in Farrell 1989 (Appendix G, pp.207-212) in the case of *produced* and *make*:

produce	*generate*	*make*	*emit*	
+		+		*noise*
+	+			*heat*
+			+	*light*
+	+		+	*energy*
+			+	*sound*
+			+	*electrons*
		+		*sthg vary*
+	+			*electricity*
		+		*a connection*
		+		*an adjustment*

5 Conclusion

It is hoped that this work has succeeded in showing the relevance of lexical research to materials development in ESP. In spite of the small scale of the investigation of the vocabulary of electronics English, some useful data has been produced. It would be interesting to see whether very different results would be found if a considerably larger corpus were studied. A complete collection of collocations for all items of interest, such as those presented in Farrell 1989 (Appendix G, pp.207-212) for *produced*, *make* and *connected*, would provide an extremely valuable body of data for subject-specific materials design. This may be most rapidly achieved, of course, where the corpus is held in computer-useable storage and the surrounding context of every occurrence of any graphic word can be instantly displayed using concordancing procedures (Sinclair 1985, p.93).

In addition, the examination of semi-technical vocabulary should make a contribution to the continuing debate concerning the role of the ESP teacher, and also provide some useful data for common core ESP course design with a lexical emphasis. A direction for future research might be a large-scale study of a number of scientific disciplines using computer facilities along the lines of the COBUILD project, which would supply vastly richer information about the general vocabulary of science than the list of graphic words in the LOB Corpus study.

Finally, it is hoped that this work will help to redress the balance a little in favour of vocabulary. At the very least, as McCarthy puts it: "It has yet to be proved that giving vocabulary at least as much weight in exercises ... as grammar ... is harmful, slows progress, or mars long-term achievement" (1984, p.14).

References

Allen, J. P. B., & H. G. Widdowson (eds.), 1974: *English in Focus: English in Physical Science*. London: Oxford University Press.

Abbs, B., & I. Freebairn, 1984: *Building Strategies*. London: Longman.

Allsop, J., 1979: *English for Cambridge First Certificate*. Eastbourne: Cassell.

Barber, C. L., 1962: "Some measurable characteristics of modern scientific prose", in Swales, pp.3-14.

Bates, M., 1978: "Writing *Nucleus*", in Mackay & Mountford, pp.78-98.

———— & T. Dudley-Evans, 1976: *Nucleus: General Science: Stu-*

dents' Book. Harlow: Longman.
Beardwood, L., H. Templeton & M. Webber, 1978: *A First Course in Technical English: Students' Book 1*. London: Heinemann.
----------, 1979: *A First Course in Technical English: Students' Book 2*. London: Heinemann.
Candlin, C., 1988: "Preface", in Carter & McCarthy, pp.vii-ix.
Carroll, J. B., P. Davies & B. Richman, 1971: *Word Frequency Book*. New York: American Heritage.
Carter, R., 1987: *Vocabulary: Applied Linguistic Perspectives*. London: Allen & Unwin.
----------, 1989: Review Articles on LDOCE and COBUILD, *International Journal of Lexicography* 2.1, pp.30-43.
---------- & M. McCarthy (eds.), 1988: *Vocabulary and Language Teaching*. London & New York: Longman.
Channell, J., 1981: "Applying semantic theory to vocabulary teaching", *ELT Journal* 35.2, pp.115-22.
Cornu, A. M., 1979: "The first step in vocabulary teaching", *Modern Language Journal* 63, pp.262-72.
Farrell, P., 1989: "Vocabulary in ESP: A Lexical Analysis of the English of Electronics and a Study of Semi-Technical Vocabulary". Unpublished M.Phil. dissertation. Dublin: Trinity College.
Fowler, W. S., 1987: *The Right Word*. Walton-on-Thames: Nelson.
Harvey, P. D., 1983: "Vocabulary learning: the use of grids", *ELT Journal* 37.3, pp.243-6.
Higgins, J., 1967: "Hard facts. Notes on teaching English to science students". Reprinted in Swales 1985, pp.30-43.
Hofland, K., & S. Johansson, 1982: *Word Frequencies in British and American English*. Harlow: Longman.
Hutchinson, T., & A. Waters, 1981: "Performance and competence in ESP", *Applied Linguistics* 2.1, pp.56-69.
----------, 1987: *English for Specific Purposes: A Learning-Centred Approach*. Cambridge: Cambridge University Press.
Ilson, R., 1985: *Dictionaries, Lexicography and Language Learning*. Oxford: Pergamon.
Inman, M., 1978: "Lexical analysis of scientific and technical prose", in Todd Trimble et al., pp.242-9.
Jackson, H., 1988: *Words and Their Meaning*. London & New York: Longman.
Johansson, S., 1975: *Some Aspects of the Vocabulary of Learned and Scientific English*. Goteborg: Acta Universitatis Gothoburgensis.
Johns, T. F., & A. Dudley-Evans, 1980: "An experiment in team-

teaching of overseas postgraduate students of transportation and plant biology". Reprinted in Swales 1985, pp.140-53.
Kennedy, C., & R. Bolitho, 1984: *English for Specific Purposes*. London: Macmillan.
Kucera, H., & W. N. Francis, 1967: *A Computational Analysis of Present-Day American English*. Providence: Brown University Press.
Leech, G., & J. Svartvik, 1975: *A Communicative Grammar of English*. London: Longman.
McCarthy, M., 1984: "A new look at vocabulary in EFL", *Applied Linguistics* 5.1, pp.12-21.
Mackay, R., & A. J. Mountford, 1978a: "A programme for post-graduate soil scientists at the University of Newcastle", in Mackay & Mountford 1978b, pp.127-60.
———— (eds.), 1978b: *English for Specific Purposes*. London: Longman.
Michea, R., 1953: "Mots fréquents et mots disponibles", *Les Langues Modernes* 47, pp.338-44.
Morais, E. (ed.), 1983: *Skills for Learning*. Walton-on-Thames: Nelson.
Morrow, K., 1980: *Skills for Reading*. Oxford: Oxford University Press
Nuttall, C., 1982: *Teaching Reading Skills in a Foreign Language*. London: Heinemann.
Praninskas, J., 1972: *American University word list*. Harlow: Longman.
Price, R. F., 1975: *A Reference Book of English Words and Phrases for Foreign Science Students*. Oxford: Pergamon.
Richards, J., 1974: "Word Lists: Problems and Prospects", *RELC Journal* 5.2, pp.69-84.
Robinson, P., 1980: *English for Specific Purposes*. Oxford: Pergamon.
Rossner, R., 1985: "The learner as lexicographer: using dictionaries in second language learning", in Ilson, pp.95-102.
Rudzka, B., J. Channell, Y. Putseys & P. Ostyn, 1981: *The Words You Need*. London: Macmillan.
Sager, J. C., D. Dungworth & P. F. McDonald, 1980: *English Special Languages*. Wiesbaden: Brandstetter Verlag.
Sinclair, J., 1985: "Lexicographic evidence", in Ilson pp.81-94.
———— & A. Renouf, 1988: "A lexical syllabus for language learning", in Carter & McCarthy, pp.140-60.
Swales, J., 1983: "Vocabulary work in LSP - a case of neglect?", *CILA Bulletin* 37, pp.21-33.

---------- (ed.), 1985: *Episodes in ESP*. Oxford: Pergamon.
Todd Trimble, M., L. Trimble & K. Drobnic (eds.), 1978: *English for Specific Purposes: Science and Technology*. Oregon: Oregon State University Press.
Trimble, L., 1985: *English for Science and Technology: A Discourse Approach*. Cambridge: Cambridge University Press.
Wallace, M., 1982: *Teaching Vocabulary*. London: Heinemann.
Webber, M., & J. Seath, 1983: *Elementary Technical English, Students' Book 1*. Walton-on-Thames: Nelson.
West, M., 1953: *A General Service List of English Words*. London: Longman.
Widdowson, H. G., 1979: *Reading and Thinking in English: Discovering Discourse*. Oxford: Oxford University Press.
----------, 1979: *Explorations in Applied Linguistics 1*. Oxford: Oxford University Press.
----------, 1983: *Learning Purpose and Language Use*. Oxford: Oxford University Press.
Willis, J., & D. Willis, 1988: *Collins COBUILD English Course*. London: Collins.
Yang, H. J., 1986: "A new technique for identifying scientific/technical terms and describing science texts", *Literary and Linguistic Computing* 1.2, pp.93-103.

APPENDIX A

LOB CORPUS SEMI-TECHNICAL WORD LIST

This is a list of graphic words defined as semi-technical, derived from the LOB Corpus Category J list (Hofland & Johansson 1982), general and technical items having been eliminated. The words are arranged in rank order of frequency and range, the first figure being the frequency and the second the range. Where a graphic word had a related lemma which occurred on the electronics English semi-technical list it is asterisked.

however*	157/66	fact	95/46	necessary*	74/38
per (cent)	145/33	level	95/29	similar*	73/44
form*	126/51	terms*	94/35	conditions*	73/34
point*	126/45	theory*	94/29	certain	72/44
results*	121/38	order	93/40	though*	72/38
field*	121/26	within*	93/38	means*	72/28
area*	119/25	group	92/22	experiments*	72/17
possible*	117/55	during*	90/42	result*	70/40
even*	113/51	great*	90/41	against*	70/35
general	111/50	way*	89/41	show*	69/39
values*	111/50	particular	88/40	process	69/28
fig.*	111/20	data	88/23	almost	68/39
therefore*	110/48	cases*	85/29	seems	68/36
thus*	109/52	influence	85/22	changes*	68/26
different*	109/43	example*	84/42	least	67/41
system*	108/35	rate*	84/29	evidence	66/32
present	104/47	model	83/12	position*	66/29
value*	104/35	rather	82/47	scale	66/19
effect*	100/40	table	82/32	function*	66/17
shown*	99/34	period	81/33	course	65/36
although*	98/48	individual	81/27	view	65/32
further	98/47	section*	81/25	required*	65/30
whole	98/44	method	80/32	length*	65/21
obtained*	98/34	surface*	80/23	temperature*	64/14
still*	97/50	(per) cent	78/19	whether	62/38
upon	97/30	type*	77/36	control*	62/16
due*	96/32	change*	76/32	various*	61/36
solution	96/21	methods	76/	problem	61/35

size	61/16	differences*	49/19	latter	43/32
end*	60/34	term*	49/19	generally	43/26
set*	60/32	place*	48/33	interest	43/26
information*	59/23	seem	48/31	stage	43/26
material*	58/27	associated	48/25	degree	43/22
act*	58/12	total	48/25	fields*	43/15
described	57/38	measured*	48/22	principle*	43/13
either*	57/36	sense	48/18	curve	43/10
shows*	57/30	side*	47/22	become*	42/29
range	57/28	mean*	47/19	factor*	42/24
functions*	57/10	areas*	47/15	law*	42/16
already	56/35	subject	46/28	action*	42/15
main	56/35	i.e.*	46/25	clear	41/29
quite	56/34	significant	46/23	provided*	41/27
analysis	56/26	groups	46/21	appears	41/26
behaviour	56/18	points*	46/21	based*	41/25
increase*	55/22	presence	46/20	just*	41/22
itself	54/28	forms*	46/19	distance*	41/15
yet	54/28	relative*	46/19	volume	41/12
knowledge	54/22	available	45/32	special	40/25
movement*	54/15	probably	45/31	likely	40/23
reason*	53/34	amount*	45/30	systems*	40/14
normal*	53/27	greater*	45/30	class	40/13
rule	53/11	equal*	45/27	suggested	39/27
particularly	53/9	series*	45/26	study*	39/26
concerned	51/31	know*	45/25	free	39/22
early	51/29	factors*	45/21	actual*	39/21
simple*	51/28	relatively	45/18	parts*	39/21
considered*	51/25	pressure	45/16	follows*	39/19
unit*	51/22	indeed	44/30	failure	39/14
constant*	51/21	single*	44/30	internal*	39/13
development	51/21	account	44/29	considerable*	38/28
average*	51/20	state	44/28	direction*	38/20
figure*	51/20	several	44/27	figures*	38/19
technique	51/14	extent	44/26	measure*	38/19
following*	50/34	common*	44/25	modern	38/19
perhaps	50/31	applied*	44/20	produced*	38/17
difference*	50/27	final	44/20	developed	38/16
complete*	50/25	apparent	44/19	design	38/14
according	50/18	structure	44/17	objects	38/10
expected	49/26	probability	44/5	consider*	37/21

e.g.*	37/19	showed	32/21	approach	28/20
reference	37/17	obtain*	32/20	followed*	28/19
earlier	36/26	approximately*	32/19	increased*	28/19
purpose*	36/22	middle*	32/18	remains*	28/19
idea	36/19	chosen	32/16	except	28/18
elements	36/17	measurements*	32/9	accepted	28/17
situation	35/22	motion*	32/9	containing*	28/16
added*	35/20	list	32/5	materials*	28/12
calculated*	35/19	becomes*	31/25	base*	28/9
relation*	35/19	certainly	31/22	output*	28/6
types*	35/18	kind	31/22	clearly	27/23
rates	35/16	next	31/22	especially	27/21
occur*	35/15	problems	31/19	possibility*	27/20
selection	35/11	natural	31/18	sufficient	27/19
towards	34/22	things	31/18	defined*	27/16
reduced*	34/20	linear	31/12	including	27/16
proportion*	34/17	experimental*	31/10	levels	27/15
investigation*	34/16	depth	31/8	mere	27/14
hence*	34/15	addition*	30/20	theoretical*	27/13
previous	34/15	observe*	30/17	reduction*	27/12
effects*	34/14	examination	30/16	space*	27/12
concentration	34/6	reported	30/14	capacity	27/11
true	33/24	shape	30/14	experiment*	27/11
rise*	33/23	numbers*	30/11	related*	26/22
effective*	33/21	properties*	30/9	became*	26/19
allowed*	33/20	called*	29/22	mainly	26/19
support	33/20	increasing*	29/22	apart	26/18
activity	33/18	mentioned	29/22	attention	26/18
comparison*	33/17	real	29/21	completely*	26/18
pattern	33/17	held*	29/19	doubt	26/18
primary*	33/16	centre*	29/18	drawn*	26/18
rapidly*	33/15	significance	29/19	finally	26/18
source*	33/15	fall*	29/15	provide*	26/18
upper*	33/11	practical*	29/15	determined*	26/17
classes	33/10	explanation*	29/14	distribution	26/16
sound*	33/8	initial	29/14	independent*	26/15
once	32/27	stages	29/12	argument	26/14
appear	32/24	agreement	29/11	cause*	26/14
basis*	32/24	secondary*	29/9	tend*	26/14
compared*	32/22	interesting	28/21	composition	26/12
limited	32/22	separate*	28/21	transfer*	26/12

event	26/9	obvious	23/19	manner	21/18
thickness	26/8	along	23/18	closely*	21/17
rules	26/7	attempt	23/17	details	21/17
speed*	26/5	impossible	23/17	directly*	21/17
indicate*	25/21	rest	23/17	indicated*	21/17
fully	25/20	assumed*	23/16	purposes*	21/17
nevertheless	25/20	direct*	23/16	require*	21/17
reached*	25/20	regard	23/16	depend*	21/16
discussion	25/19	remain*	23/16	hardly	21/16
discussed	25/18	equivalent	23/15	instance	21/16
corresponding	25/17	expression	23/15	maximum*	21/16
affected*	25/16	produce*	23/15	respectively	21/16
procedure	25/16	techniques	23/15	simply*	21/16
extended	25/15	received*	23/14	apparatus*	21/15
supply*	25/14	represented*	23/14	exactly	21/15
research	25/12	amounts*	23/12	refer	21/14
concept	25/11	product*	23/12	accuracy*	21/14
divided	25/11	central*	23/10	grounds	21/13
sample	25/11	concepts	23/9	highly	21/13
measurement*	25/6	spread	23/8	reasons*	21/13
throughout	24/20	flow*	23/7	weight	21/13
marked	24/19	survey	23/5	hypothesis	21/9
whereas	24/19	noted*	22/19	scientific	21/9
considering*	24/18	introduction	22/18	subjects	21/7
nearly*	24/18	appropriate	22/17	no.	21/6
sufficiently	24/18	expressed	22/17	absence	20/18
satisfactory	24/17	relevant	22/17	examples*	20/17
standard	24/16	suggests	22/17	recent	20/17
included	24/14	detail	22/16	removed*	20/17
definition*	24/12	slightly*	22/16	similarly*	20/17
external	24/12	appeared	22/15	circumstances	20/16
occurs*	24/12	etc.	22/15	determine*	20/15
height	24/10	merely	22/15	placed*	20/15
mechanism	24/10	unless	22/15	immediately	20/14
occurrence	24/10	application	22/14	strength*	20/14
liquid	24/8	recognized	22/14	condition*	20/13
programme	24/7	property*	22/12	variation*	20/13
solid	24/6	potential	22/11	consequently	20/12
load	24/5	rapid*	22/11	limits	20/12
established	23/20	component*	22/7	operation*	20/12
nor	23/19	performance*	22/6	relations*	20/12

features	20/11	isolated	20/8	objective	20/6
growth	20/11	temperatures	20/8	velocity*	20/6
object	20/11	division	20/7	displacement	20/5
occurred*	20/11	thermal	20/7	place*	20/5
uniform*	20/10	interpretation	20/6		

APPENDIX B

ELECTRONICS ENGLISH ALPHABETICAL LIST

This is an alphabetically arranged list of all graphic words which were grouped under a lemma with a total occurrence of over 5 in the electronics English corpus. The figure refers to the rank order position of the lemma entry in Appendix C under which the graphic word has been included.

a	3	actual	149	altering	204
ability	227	actually	149	alternate	76
able	227	add	203	alternately	76
absolute	229	added	203	alternating	76
absolutely	229	adding	203	alternative	76
about	126	addition	203	although	180
above	143	additional	203	always	204
a.c.	63	aerial	194	ammeter	117
accelerate	213	affect	205	ammeters	117
accelerated	213	affected	205	amount	195
accelerating	213	after	217	amounts	195
acceleration	213	again	222	amp	204
accumulator	207	against	227	ampere	204
accumulators	207	air	191	amperes	204
accuracy	218	all	147	amps	204
accurate	218	allow	197	amplification	66
accurately	218	allowance	197	amplified	66
across	51	allowing	197	amplifier	66
act	149	allows	197	amplifiers	66
acting	149	also	118	amplify	66
action	149	alter	204	amplifying	66
acts	149	altered	204	amplitude	154

APPENDIX C

ELECTRONICS CONDENSED RANK FREQUENCY LIST

This is a list of all lemmas representing a total occurrence of more than 5 tokens in the electronics English corpus, arranged in rank order of frequency and range. The first figure following the lemma is the frequency and the second the range. Lemmas with the same frequency and range share a rank position.

1. the	2034/10	32. which	89/10	58. transistor	56/2
2. is	879/10	33. or	86/10	59. time	53/8
3. a	850/10	34. potential	86/7	60. wave	53/7
4. of	692/10	35. you	81/9	61. flow	52/7
5. to	439/10	36. shown	80/10	62. varying	51/10
6. and	394/10	37. produced	79/8	63. a.c.	50/7
7. in	367/10	38. energy	79/6	64. see	49/9
8. current	284/10	plate	79/6	65. then	48/10
9. it	207/10	39. electric	78/10	66. amplifier	48/6
10. fig.	207/7	make	78/10	67. through	47/10
11. this	188/10	40. wire	77/8	68. negative	47/8
12. voltage	161/9	41. capacitor	77/6	69. switch	47/6
13. by	149/10	42. moving	75/8	70. experiment	46/7
14. as	145/10	43. not	70/10	71. emitter	46/6
15. circuit	140/10	will	70/10	72. collector	46/2
16. with	134/10	44. when	69/10	73. two	45/9
17. on	121/10	45. connected	68/10	74. between	45/8
18. used	119/10	46. frequency	68/7	do	45/8
19. from	117/10	47. change	65/10	75. beam	45/3
20. that	112/9	so	65/10	76. alternating	44/7
21. resistance	112/6	48. oscillator	64/7	77. given	43/10
22. charge	111/9	49. measure	63/8	they	43/10
23. has	109/10	50. supply	62/10	78. more	42/9
24. electron	108/7	value	62/10	79. high	42/8
25. can	104/10	51. across	62/7	80. force	41/5
26. for	98/10	52. one	60/10	81. e.m.f.	40/6
27. at	97/10	53. output	60/9	82. anode	40/4
28. coil	97/6	54. we	60/8	83. small	39/10
29. magnetic	92/8	55. base	60/5	84. effect	39/9
30. if	90/10	56. direction	59/8	increased	39/9
31. field	90/8	57. positive	58/8	line	39/9

low	39/9	117. ammeter	26/5	144. because	19/8
cathode	39/9	118. also	25/10	how	19/8
85. screen	38/4	119. graph	25/7	takes	19/8
86. same	38/10	may	25/7	145. must	19/7
87. only	37/10	would	25/7	146. transmitter	19/4
88. what	37/9	120. control	25/6	147. all	18/8
89. called	36/10	121. passes	24/8	148. form	18/7
90. diode	35/6	result	24/8	example	18/7
length	35/6	zero	24/8	however	18/7
91. radio	35/3	122. needed	24/7	149. acts	18/6
92. now	34/8	should	24/7	case	18/6
93. heat	34/7	123. found	24/6	down	18/6
94. deflection	34/3	closed	24/6	gain	18/6
95. difference	33/8	124. calculate	24/5	opposite	18/6
96. applied	33/7	set	24/5	150. peak	18/3
97. metal	33/6	125. source	23/8	151. velocity	18/2
98. induced	33/3	126. about	23/7	152. cause	17/7
99. balance	33/1	junction	23/7	could	17/7
100. terminal	32/7	127. tube	23/5	153. common	17/4
101. filament	32/4	128. core	23/2	154. amplitude	17/3
102. large	31/9	129. battery	22/7	155. silicon	17/2
103. light	31/6	power	22/7	156. voltmeter	16/7
104. but	30/9	130. conductor	22/6	157. thus	16/6
105. p.d.	30/7	note	22/6	where	16/6
turn	30/7	since	22/6	158. transfer	16/3
type	30/7	131. temperature	22/5	159. error	16/2
106. cell	30/5	132. stored	22/4	mass	16/2
107. horizontal	29/6	133. vertical	22/3	path	16/2
108. other	28/9	134. C.R.O.	22/2	160. modulated	16/1
point	28/9	135. simple	21/8	161. such	15/8
109. d.c.	28/7	136. number	21/7	162. around	15/7
thermionic	28/7	there	21/7	greater	15/7
110. cm.	28/3	137. obtain	21/6	possible	15/7
111. some	27/9	138. ray	21/4	similar	15/7
112. very	27/8	signal	21/4	than	15/7
113. constant	27/7	139. system	21/3	163. component	15/6
each	27/7	140. open	20/6	end	15/6
114. carrier	27/6	141. reading	20/5	following	15/6
115. solenoid	27/1	142. sphere	20/1	happens	15/6
spot	27/1	143. above	19/10	material	15/6
116. up	26/8			164. maximum	15/5

75

steady 15/5
165. pole 15/4
rises 15/4
166. feedback 15/2
167. insulated 14/8
168. first 14/7
reverse 14/7
second 14/7
169. device 14/5
fails 14/5
rectifier 14/5
work 14/5
170. strength 14/4
171. block 14/2
grid 14/2
reactance 14/2
receiver 14/2
transformer 14/2
uniform 14/2
172. both 13/9
173. consists 13/8
174. any 13/7
required 13/7
175. continuously 13/6
less 13/6
operate 13/6
part 13/6
straight 13/6
176. ohm 13/5
ratio 13/5
177. wind 13/4
178. held 13/3
iron 13/3
load 13/3
semiconductor 13/3
sound 13/3
179. spring 13/1
180. depends 12/7
still 12/7
though 12/7
way 12/7
181. complete 12/6
consider 12/6
keep 12/6
means 12/6
reduced 12/6
182. easily 12/5
generator 12/5
until 12/5
183. speed 12/4
electromagnetic 12/4
184. earth 12/3
near 12/3
185. photo 12/2
protective 12/2
186. radius 12/1
187. over 11/8
188. another 11/7
189. even 11/6
indicated 11/6
piece 11/6
190. back 11/5
illustrated 11/5
lead 11/5
plotting 11/5
repeat 11/5
191. air 11/4
drop 11/4
192. flux 11/3
193. termed 11/2
194. aerial 11/1
relay 11/1
195. amount 10/7
compare 10/7
suitable 10/7
usually 10/7
196. bulb 10/6
placed 10/6
197. allow 10/5
attracted 10/5
important 10/5
T.V. 10/5
198. decreases 10/4
rapidly 10/4
remove 10/4
199. copper 10/3
hence 10/3
lost 10/3
position 10/3
valve 10/3
200. armature 10/2
compass 10/2
demonstration 10/2
201. gate 10/1
202. together 9/7
203. added 9/6
before 9/6
just 9/6
occur 9/6
said 9/6
204. alter 9/5
always 9/5
amp 9/5
brightness 9/5
205. affect 9/4
i.e. 9/4
meter 9/4
parallel 9/4
push 9/4
represents 9/4
206. distance 9/3
galvanometer 9/3
off 9/3
series 9/3
tends 9/3
207. accumulator 9/2
contact 9/2
picture 9/2
test 9/2

208. gun 9/1
209. many 8/7
210. much 8/6
why 8/6
211. paper 8/5
red 8/5
therefore 8/5
212. becomes 8/4
detect 8/4
fed 8/4
law 8/4
mid-point 8/4
normal 8/4
provide 8/4
reach 8/4
round 8/4
side 8/4
213. acceleration 8/3
assuming 8/3
construction 8/3
divider 8/3
instruction 8/3
slow 8/3
214. chip 8/2
electroscope 8/2
fluorescent 8/2
radiation 8/2
rheostat 8/2
space 8/2
215. few 7/6
principle 7/6
216. best 7/5
enough 7/5
separate 7/5
slightly 7/5
217. after 7/4
build 7/4
drive 7/4
equal 7/4
explained 7/4
independent 7/4
page 7/4
prevent 7/4
rate 7/4
region 7/4
study 7/4
three 7/4
unit 7/4
218. accurate 7/3
coated 7/3
cut 7/3
drawn 7/3
intensity 7/3
often 7/3
put 7/3
surface 7/3
219. angle 7/2
bias 7/2
enters 7/2
equation 7/2
information 7/2
oxide 7/2
shorting 7/2
visible 7/2
220. drain 7/1
escape 7/1
flywheel 7/1
221. condition 6/5
either 6/5
essential 6/5
property 6/5
pulled 6/5
222. again 6/4
approximate 6/4
atom 6/4
below 6/4
centre 6/4
come 6/4
difficult 6/4
during 6/4
goes 6/4
practical 6/4
quantity 6/4
replaced 6/4
single 6/4
write 6/4
223. apparatus 6/3
characteristic 6/3
conventional 6/3
every 6/3
factor 6/3
internal 6/3
investigate 6/3
leave 6/3
narrow 6/3
purpose 6/3
relation 6/3
symbol 6/3
224. contains 6/2
correction 6/2
coulomb 6/2
discharge 6/2
due 6/2
little 6/2
secondary 6/2
too 6/2
225. bath 6/1
blade 6/1
integrated 6/1
impedance 6/1
jockey 6/1
perpendicular 6/1
soft 6/1
ultra-violet 6/1
226. diagram 5/5
known 5/5
remains 5/5
under 5/5

227. able	5/4	molecule	5/3	width	5/2
against	5/4	perform	5/3	zinc	5/2
area	5/4	repulsion	5/3	230. audio	5/1
cross	5/4	229. absolute	5/2	channel	5/1
e.g.	5/4	average	5/2	dielectric	5/1
here	5/4	black	5/2	distortion	5/1
observed	5/4	crystal	5/2	earphone	5/1
proportional	5/4	determine	5/2	hydrogen	5/1
rotary	5/4	flies	5/2	mode	5/1
theory	5/4	ion	5/2	neon	5/1
try	5/4	maintained	5/2	primary	5/1
228. defined	5/3	metre	5/2	razor	5/1
doped	5/3	needle	5/2	r.m.s.	5/1
fixed	5/3	pin	5/2	send	5/1
function	5/3	pivot	5/2	sinusoidal	5/1
get	5/3	reason	5/2	spark	5/1
gradually	5/3	section	5/2	traced	5/1
half	5/3	slide	5/2	triode	5/1
instrument	5/3	summarised	5/2		

APPENDIX D

ELECTRONICS ENGLISH GENERAL WORDS

This appendix divides the list presented in Appendix C into three separate lists of general, semi-technical, and technical lemmas. Each list is arranged in rank order of frequency and range. In the case of the semi-technical list, all lemmas representing graphic words which occurred on the LOB Corpus semi-technical list are asterisked.

GENERAL WORDS

the	in	on	for
is	it	used	at
a	this	from	if
of	by	that	which
to	as	has	or
and	with	can	you

make	light	first	after
moving	but	second	page
not	other	strength	three
will	some	both	cut
when	very	any	often
so	each	keep	put
across	up	means	again
one	may	easily	below
we	would	until	come
time	passes	speed	difficult
see	zero	another	goes
then	needed	air	write
through	should	important	every
two	found	T.V.	leave
between	closed	together	correction
do	about	before	little
given	there	always	too
they	open	off	soft
more	above	picture	known
small	because	test	able
line	how	many	cross
same	takes	much	here
only	must	why	try
what	all	paper	fixed
called	down	red	get
length	could	round	half
radio	where	slow	black
now	around	few	width
heat	than	best	send
large	happens	enough	

ELECTRONICS ENGLISH SEMI-TECHNICAL WORDS
(lemmas also on LOB semi-technical list asterisked)

fig.*	supply*	alternating	difference*
shown*	output*	high	metal
produced*	direction*	effect*	balance
connected	flow*	increased*	turn
change*	varying*	low	type*
measure*	emitter	deflection	horizontal

constant*	possible*	termed*	instruction
carrier	similar*	amount*	space*
spot	component*	compare*	principle*
also	end*	suitable*	separate*
control*	following*	usually	slightly*
result*	material*	placed*	build
calculate*	maximum*	allow*	drive
set*	steady	attracted	equal*
source*	rises*	decreases	explained*
junction	reverse	rapidly*	independent*
tube	device	remove*	prevent
core	falls*	hence*	rate*
note*	received	lost	region
since	uniform*	position*	study*
temperature*	consists of	demonstration	unit*
stored	required*	added*	accurate*
vertical	continuously	just*	covered
simple*	less	occur*	drawn*
number*	operate	said	intensity
obtain*	part	alter	surface*
system*	straight	brightness	angle
reading	wind	affect*	enters
sphere	held	i.e.*	information*
transmitter	sound*	parallel	shorting
form*	spring	represents*	visible
example*	depends on*	distance*	escape
however*	still*	tends*	condition*
acts*	though*	contact	either*
case*	way*	therefore*	essential
gain	complete*	becomes*	property*
opposite	consider*	detect	pulled
peak	reduced*	fed	approximate*
velocity*	near*	law*	centre*
cause*	protective*	mid-point*	during*
common*	over	normal*	practical*
thus*	even*	provide*	quantity
transfer*	indicated*	reach*	replaced
error	piece	side*	single*
path	illustrated	acceleration	apparatus*
such	repeat	assuming*	conventional
greater*	drop	construction	internal*

investigate*	diagram	theory*	determine*
narrow	remains*	defined*	files
purpose*	under	function*	maintained
relation*	against*	gradually	reason*
symbol	area*	instrument	section*
contains*	e.g.*	perform*	slide
due to*	observed*	repulsion	summarised
secondary*	proportional*	absolute	distortion
perpendicular	rotary	average*	primary*

ELECTRONICS ENGLISH TECHNICAL WORDS

current	collector	amplitude	plotting
voltage	beam	silicon	flux
circuit	force	voltmeter	aerial
resistance	e.m.f.	mass	relay
charge	anode	modulated	bulb
electron	cathode	pole	copper
coil	screen	feedback	valve
magnetic	diode	insulated	armature
field	applied	rectifier	compass
potential	induced	work	gate
energy	terminal	block	amp
plate	filament	grid	meter
electric	p.d.	reactance	push
wire	cell	transformer	galvanometer
capacitor	point	ohm	series
frequency	d.c.	ratio	accumulator
oscillator	thermionic	iron	gun
value	cm.	load	divider
base	solenoid	semiconductor	chip
positive	ammeter	generator	electroscope
transistor	graph	electromag-	fluorescent
wave	battery	netic	radiation
a.c.	power	earth	rheostat
amplifier	conductor	photo	bias
negative	C.R.O.	radius	equation
switch	ray	back	oxide
experiment	signal	lead	drain

flywheel	impedance	pin	neon
atom	jockey	pivot	razor
characteristic	ultra-violet	zinc	r.m.s.
factor	doped	audio	sinusoidal
coulomb	molecule	channel	spark
discharge	crystal	dielectric	traced out
bath	ion	earphone	triode
blade	metre	hydrogen	
integrated	needle	mode	

APPENDIX E

THE ELECTRONICS ENGLISH CORPUS

1. Barker, B. G., 1981: *Foundation Electronics* (London: Nelson), pp.89-91, 2000 words.
2. Bishop, O. N., 1981: *Physics: A Practical Approach* (London: Macmillan), pp.324-7, 2025 words.
3. Sladdin, M., 1983: *Elementary Electronics* (London: Hodder & Stoughton), pp.72-6, 2032 words.
4. Nelkon, M., & J. M. Ogborn, 1978: *Advanced Level Practical Physics* (London: Heinemann), pp.146-53, 2098 words.
5. Nelkon, M., & P. Parker, 1982: *Advanced Level Physics* (London: Heinemann), pp.801-10, 2000 words.
6. Davies, D., & D. McCullough, 1981: *Physics for First Examinations* (Edinburgh: Holmes McDougall), pp.200-5, 2000 words.
7. Jardine, J., 1972: *Physics is Fun* (London: Heinemann), pp.78-83, 2000 words.
8. Webster, M., 1983: *Essentials of Higher Physics* (London: Heinemann), pp.87-94, 2000 words.
9. Pople, S., 1982: *Explaining Physics* (Oxford: Oxford University Press), pp.312-6, 2009 words.
10. Jardine, J., 1983: *Higher Physics* (London: Heinemann), pp. 146-50, 1853 words.

Spring 1986

4. D. G. Little and A. J. Grant, *Learning German without a teacher. Report on a self-instructional programme for undergraduate students of Engineering Science at Trinity College, Dublin, 1982-1984* (60pp.)
15. Máire Owens, *Eithne: a study of second language development* (62pp.)
16. D. Wilson and D. Sperber, *Pragmatics: an overview* (36pp.)

Autumn 1986

17. Aibhe Ni Chasaide& Eugene Davis, *A data-processing system for quantitative analysis in speech production* (28pp.)
18. Seán M. Devitt, *Learning a foreign language through the media* (69pp.)
19. Meriel Bloor & Thomas Bloor, *Languages for specific purposes: practice and theory* (34pp.)

Spring 1988

20. D. G. Little & D. M. Singleton, *Authentic materials and the role of fixed support in language teaching: towards a manual for language learners* (26pp.)

Spring 1989

21. Seán M. Devitt, *Classroom discourse: its nature and its potential for language learning* (72pp.)
22. V. J. Cook, *The relevance of grammar in the applied linguistics of language teaching* (43pp.)

Spring 1990

23. Sera De Vriendt & Pete Van de Craen, *Bilingualism in Belgium: a history and an appraisal* (52pp.)
24. David Singleton, *The cross-linguistic factor in second language learning: a report on some small scale studies recently conducted at the CLCS* (20pp.)

Autumn 1990

25. Paul Farrell, *Vocabulary in ESP: a lexical analysis of the English of electronics and a study of semi-technical vocabulary* (82pp.)
26. David Singleton, *The TCD Modern Languages Research Project: objectives, instruments and preliminary results* (19pp.)
27. Federica Scarpa, *Contrastive analysis and second language learners' errors: an analysis of C test data elicited from beginners in Italian* (47pp.)